JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 30 (2025), pp. 25–60

doi:10.5311/JOSIS.2025.30.385

RESEARCH ARTICLE

Automated validation of route instructions in indoor environments

Reza Arabsheibani, Stephan Winter, and Martin Tomko

Department of Infrastructure Engineering, The University of Melbourne, Australia

Received: June 27, 2024; returned: October 18, 2024; revised: November 4, 2024; accepted: January 11, 2025.

Abstract: Indoor wayfinders often rely on verbal route directions, particularly in situations where other navigational aids may be unavailable or less effective. Ensuring the clarity and validity of these instructions is particularly important for navigation in complex indoor environments, such as airports and malls. However, current methods lack a reliable, systematic approach to computationally ensuring the *a-priori* validity of route instructions, failing to provide certainty to agents that they will be able to follow instructions successfully. Here we show a novel computational model for validating indoor route instructions, applicable to a wide range of indoor environments and turn-based grammars. Using a synthetic dataset of indoor floorplans with varying complexities, we demonstrate the model's capability to validate route instructions systematically. We systematize the requirements for route instruction validation in the framework which assesses instructions based on understandability, executability, path-following, and destination guidance. Our findings highlight the effectiveness of nuanced grammars, such as 8-sector grammar, for complex layouts and confirm the applicability of simpler grammars, like 4-sector grammar, for right-angle constrained environments. Importantly, we identify a transition point where the benefits of increased grammatical complexity on the descriptions of the turns are no longer productively supporting a reduction in turn ambiguity in the environments. This research shifts the field from subjective, time-consuming human evaluations to a computational approach, enhancing the reliability of indoor navigation systems.

Keywords: route instructions, validation, floorplan complexity, grammar, instruction-following agents, text-based games

1 Introduction

Indoor wayfinders often rely on verbal route directions, particularly when other navigational aids such as maps or signage are unavailable or insufficient. This is especially true in environments like airports, museums, or large shopping malls, where clear and valid verbal instructions can be critical for effective navigation [12, 23, 28]. For instance, a visually impaired individual navigating a busy airport can benefit from concise and understandable verbal instructions [70]. Verbal route instructions are common navigation aids for wayfinding in unfamiliar environments, especially indoors, where other wayfinding assistance methods may be unavailable or less effective [13, 42, 67]. Route instructions describe decision points and guide the wayfinder (the person or agent following the instructions) on actions to take [30] to reach the destination. Route instructions complement information perceived directly from the environment, past experiences, or other spatial sources (e.g., you-are-here maps [36]). The validation of route instructions targets the understanding [60] and execution [9] capabilities of the agent concerning the navigation instructions.

Despite their critical role in indoor navigation, we lack an automated validation framework for route instructions. Various types of uncertainties may affect the validity of route instructions [61]. Wihtout validation, route instructions are provided without *explicit* certainty that an agent with given skills may successfully follow them and reach the intended destination. The absence of validation is critical in complex indoor environments, where uncertainties can arise from the mismatch between the formulation of instructions (using a specific *grammar*), and the complexity of the decision points where the instructions are interpreted and matched to the environment. Therefore, while in a simple environment a wayfinder may be able to navigate using a specific grammar, they may be challenged in finding their destination when using the same grammar in a more complex environment [3]. Grammars proposed in the literature are meant for *generating* instructions [11, 35, 46, 48, 54, 55, 69] rather than following them. Consequently, it remains unclear how different grammars contribute to successful navigation, and therefore how they lead to *valid* route instructions across various indoor complexities.

The research question addressed in this study is "*How to develop a computational model for the validation of route instructions across various indoor environments and grammars?*". To answer this question, we first propose four criteria for evaluating *validity of route instructions*: the route instructions must be (1) *understandable*, (2) *executable*, (3) *specifying a path*, and (4) leading to the *intended destination*. We treat route instructions as a sequence of individual turn directions. The validity of route instructions is therefore limited by the validity of its constituent turn directions. In this study we focus on turn-based grammars (such as "*go sharp right*"), because they are ranked as the most critical type of navigational information according to participant preferences [10].

We hypothesize that the validity of turn-based route instructions is inversely correlated with floorplan complexity, with more nuanced grammars (e.g., 8-sector) demonstrating higher navigation success rates in complex environments than simpler grammars (e.g., 4-sector). However, there is an optimal level of grammatical complexity, beyond which additional detail may lead to reduced effectiveness.

This hypothesis entails that as floorplan complexity increases, the route instructions' validity decreases more significantly for simpler grammars. In contrast, more complex grammars maintain high validity to some extent but may also experience a decline in performance at very high levels of grammatical complexity. By testing this framework on a dataset with varying complexities, we aim to demonstrate its applicability.

The main contributions of the present study are:

1. Proposing an automated model for validating indoor route instructions that can be applied to a wide range of indoor environments;

2. Systematizing the requirements for the validity of route instructions;
3. Developing a new synthetic dataset of indoor floorplans, designed to cover and exceed the spectrum of complexities of real-world floorplans, to systematically assess the impact of floorplan complexity on the validity of route instructions.

Most existing approaches to validate route directions rely on human participants, which is time-consuming, subjective [69], and not applicable upon instruction generation. Despite much prior work on generating route instructions and studying human comprehension, we lack a systematic means to computationally validate whether an agent can successfully follow a set of route instructions to the intended destination. Our research aims to develop a universal computational model for validating route instructions for a diverse combination of environment and grammar formalism used to express the instructions. Such a model would enable automated validation of route instruction upfront, before deployment to users, or in navigation systems for specific indoor environments. Such a model would also allow for systematic analysis of how different grammar and environmental complexities impact the likelihood of encountering invalid instructions.

In this study, we focus on validating verbal route instructions for indoor navigation rather than generating such instructions. We only assume that these instructions were generated following a consistent grammar for expressing turn actions. We then computationally assess whether a given set of such turn-based route instructions can be successfully executed (i.e., understood and followed) to reach the intended destination by an agent *capable of understanding* the grammar used to generate these instructions. To achieve this, we interpret the instructions strictly, ensuring that the interpretation in the execution step aligns with the generation process to maintain a consistent understanding of the instructions.

The approach presented here can be applied to environments that meet specific, but sufficiently general criteria. Specifically, it can be applied to 2D indoor environments with no vertical connectivity (since the considered turn-grammars use only turns in the horizontal plane and do not consider up and down) and with some structure (for turns to make sense). The latter excludes environments with large unstructured spaces or those featuring fuzzy transitions (e.g., spaces with furniture or curved spaces). Here, the validation strictly relies on turn-based instructions, and these instructions must match the grammar known to the agents. The agent does not account for prior navigation experience or individual cognitive variations, such as memory or learning effects, which may influence real-world wayfinding performance. Another limitation is that, while our model addresses turn-based route instructions, it does not account for signage, perceptual factors, or landmark-assisted navigation. Specifically, we do not incorporate the visibility or identification of landmarks, nor the potential for confusion when multiple visually similar objects are present.

Wayfinders that need to navigate regularly in unknown or partially unknown environments, such as tourists, shoppers in large malls, or wayfinders with specific abilities, will benefit from the ability to pre-validate route instructions contextualized to the complexity of the environment to be navigated. Additionally, operators of these environments can ensure that their instruction systems are well-matched to the complexities of their spaces. This research may also inform the design of verbal interfaces with embodied robotic assistants in such environments (delivery robots, intelligent wheelchairs).

The rest of the paper is structured as follows. In Section 2, we discuss the background and related work, highlighting the gaps in current validation approaches. Section 3 introduces our universal validation model, detailing the validation criteria, environment simulation, and data processing methods. Section 4 demonstrates the application of our model using a synthetic dataset of indoor floorplans, focusing on the impact of floorplan complexity on route instruction validity. In Section 4.5, we present the implementation of our framework, followed by the results and discussion in Section 5, where we evaluate the performance of different grammars in varying floorplan complexities. Finally, Section 6 concludes the study and suggests future research directions.

2 Background

2.1 Validation of route instructions

Tomko et al. [61] elaborated on the concept of *uncertainty* in route instructions, highlighting that instructions may be invalid or contain uncertainties that can impede successful navigation. Such uncertainties can arise from various factors, including conflicting information or incomplete directions, which may prevent the wayfinder from making correct decisions or reaching the intended destination. Our computational validation method focuses on assessing the validity of route instructions based on predefined criteria to ensure their reliability and effectiveness. Specifically, uncertainties can be categorized into missing information, ambiguity arising from a mismatch between the grammar of route instructions and the geometry of the environment, and errors. In this paper, we primarily address ambiguity due to its direct relevance to our validation approach.

The pragmatic validity of route instructions is determined by their ability to successfully guide followers to their intended destination [22,46,62]. Past research predominantly evaluated route instructions through human subject experiments and subjective assessments, lacking systematic, computational validation approaches. An early study sought to identify rules for clarifying route instructions [57], while others used expert opinions and human subjects to evaluate the *quality* of instructions [47]. Human subjects are investigated for comparing *at-once* route instruction generation and the complexity of dialogue-based *in-situ* route instructions [26,58]. While insightful, these human-centric validation techniques are extremely time- and resource-intensive to conduct at scale.

Tschander et al. [64] proposed a formal framework for simulating an agent following route instructions, bridging linguistic inputs with spatial navigation. Similar to our approach, they deliberately abstracted away from perceptual challenges (like object recognition and re-identification), acknowledging these as open problems in computer vision, to focus purely on the geometric aspects of instruction following. Their model centers on simulating the interpretation and execution of instructions, while in contrast, here we focus on the validation of these instructions.

Navigation research has extensively explored the role of landmarks in helping individuals find their way [44,66]. Landmarks act as prominent reference points in an environment, making it easier for people to orient themselves and navigate from one place to another. In the domain of Vision and Language Navigation (VLN), some models have been developed to integrate route instructions with corresponding navigational trajectories, aiming to improve the alignment between verbal directions and physical movement [69]. However, these models often focus on textual evaluation metrics—such as how closely generated instructions match human descriptions—which may not fully capture the real-world complexities of navigation, like environmental changes or individual user preferences. Additionally, while there are guidelines and evaluation methods proposed for navigation systems that assist people with disabilities [68], these approaches usually address a limited set of predefined situations. This narrow focus means that many potential applications, especially those requiring adaptability to diverse scenarios and user needs, remain underexplored.

Usability studies provide insights into adaptive indoor wayfinding systems [16, 17], but evaluations predominantly emphasize cognitive load reduction and environmental awareness enhancement, sidestepping comprehensive route instruction validation. Another study for examining comfort and confusion perception during wayfinding in indoor environments is presented [65], indicating a potential avenue for the development of robust route planning algorithms. The direct translation of these findings into validation methodologies remains, however, an ongoing challenge. Thus, while existing studies contribute valuable insights, there is a pressing need for holistic route instructions validation frameworks.

Several previous works have focused on enhancing the quality of route instructions by addressing different aspects of navigational complexity. For example, the concept of *"simplest paths,"* is introduced [18] where the primary goal is to reduce cognitive load by identifying routes that are (cognitively) simpler, instead of merely being shortest. This approach minimizes the number of complex decisions or turns, thus resulting in routes that are presumably also easier to execute, although the path length might be longer. Similarly, Haque et al. [32] developed an algorithm that prioritizes minimizing turn ambiguities at complex intersections, proposing reliable routes that balance distance and navigational complexity. These reliable paths require simpler instructions like "veer right" rather than just "turn right," reducing the likelihood of navigational errors. Building on these foundational ideas, subsequent works, e.g., like [40], explored ways to reduce navigational complexity by focusing on minimizing the number of crossings and decision points along the route. Additionally, the concept of a *"most recoverable route,"* was proposed [4] to support wayfinders in recovering from errors, thereby improving wayfinding robustness. These studies share a common focus on improving the usability of route instructions through computational methods that account for environmental features and cognitive simplicity, yet they differ from our approach, which focuses on replaying and executing route instructions in situ with an agent that follows well-defined grammatical structures.

2.2 Route instruction grammars

A route is described to a wayfinder in route instructions integrating spatial, linguistic, and contextual knowledge elements of distinct types, summarily constituting a *grammar*. The ability to comprehend a grammar is crucial for the generation of route instructions (i.e., representing the route) as well as for the understanding and following (and thus also validating) route instructions.

The correct interpretation of route instructions depends on the specified grammar and context, and can result in a rich descriptive specifications:

- Turn-based grammars express route instructions as a sequence of atomic *turn directions*. For example, in a formal interpretation of the 4-sector model, "turn left" could mean any turn between 225 and 315 degrees. However, just from the instruction it can be challenging for the recipient to determine the exact direction without knowing which sector model has been applied by the system. Common (aligned with European languages) are the 4-sector grammar, and the 8-sector grammar. Seafarers go beyond and have a 16-sector grammar (North, North-North-West, North-West, etc.).

- Landmark-based grammars extend turn-based grammars with landmark references, used for the identification of decision points [39]. For example, “at the intersection with the Post office, turn left" expects a continued movement until an intersection with a *Post office* is reached. Landmarks can also be used for confirmation *en-route* (“pass the 7-Eleven"), or for global orientation (“head to the city centre"), adding complexity to the grammar. Route instructions can be contextualized with limiting the use of landmarks to the pragmatically necessary number [24].
- Grammars can also include distance, quantitatively and qualitatively, with references to ordinal sequences, time, and landmarks (e.g., *“at the second intersection”*, *“after that, turn left”, or “walk until you reach a T intersection”*.
- Grammars may include specifications of geometric quantities—angles in degrees, distances in metres, time in hours, minutes and seconds—although these are uncommon when the addressee is a person, but are of high utility in instructions executable by machines, e.g., robots.

Research into the structure of route instructions has identified several hierarchical sublanguages that help classify different types of directives. According to Pappu et al. [54], route instructions can be divided into four main categories: **imperatives**, which issue direct commands; **advisory**, which offer suggestions or recommendations; **grounding**, which provide context or clarification; and **meta-comment**, which include remarks about the instructions themselves. Building on this framework, Zhao et al. [69] further refined instruction grammars by categorizing them into three distinct types: *move*, *orient*, and *object*. The *move* category includes verbs that direct movement, such as "turn" or "continue." The *orient* category involves instructions related to positioning or alignment, like "face north" or "align with the building." The *object* category pertains to references to specific landmarks or objects, using verbs like "reach" or "pass by." Additionally, Zhao et al. [69] suggested a set of accepted aliases and prepositions to standardize these instructions, enhancing their clarity and consistency. To provide a comprehensive overview of the various spatial descriptions and grammatical structures used in route instructions, we have summarized the findings from these studies in Table 1. This table highlights the different categories and their corresponding linguistic elements, offering a clear reference for understanding how route instructions are formulated and evaluated in navigation research.

Table 1: Route instruction grammar models

Model	No. of classes	No. of basic actions
Skeletal description [48]	5	not defined
Wayfinding choremes [35]	4	8
Navigation primitives [11]	1	15
Walk the talk [46]	1	4
Abstract instructions [55]	4	27
Navigation taxonomy [54]	4	18
Instruction templates [69]	3	12

The influence of language on the conceptualization of turn directions has been investigated [37]. Linguistic tasks involve verbal labelling, while nonlinguistic tasks involve grouping turns based on similarity without verbal labels. We have included the cognitively-grounded grammar (CGG) in our experiments [38]. This cognitively-based grammar is based on human interpretations of turn directions has been used in research that need cognitive representation of directions on decision points [45]. The abbreviation *"CGG"* is introduced by the authors to refer to the grammar framework used in this study. Another common trait of human-generated route instructions, the generation of hierarchical route instructions is called *spatial chunking* [38], referring to the provision of linguistic aliases to a set of coarser instructions generated by rules and using symbols of a context-free grammar. For example, instead of providing repetitive instructions like *"go straight, go straight, go straight, turn left"*, spatial chunking enables composing concise instruction, e.g., *"continue straight for three blocks, then turn left."*

2.3 Floorplan complexity

Numerous studies have tried to quantify characteristics of indoor spaces that affect wayfinding, i.e., the *complexity* of the environment where wayfinding occurs. Indoor complexity can be analyzed by using metric or topological measures at two levels of detail—globally, capturing the complexity of the entire environmental layout, or locally, evaluating the complexity of its constituent parts, e.g., individual intersections.

The overall complexity of an entire floor plan layout has been assessed by measures such as the number of choices along a route, a measure that was reported to capture environmental complexity already by [8]. The complexity of floorplans has been quantified by the measure of Inter-Connection Density (ICD), corresponding to the average number of possible decisions per decision point in the environment. The measure of *intelligibility* for ordering the space based on the geometric properties and inter-objects successive connections was introduced [31].

The distinctions between how architects and non-architects perceive the relationships between building layouts and wayfinding complexity have been investigated, revealing that the judgments of groups correlate on environmental measures, including the existence of loops, the number of decision points, ICD, and shape convexity [15, 33]. The impact of different circular patterns on the ease of wayfinding in buildings has been explored [51]. The influence of individual differences and building characteristics on the complexity of indoor wayfinding tasks has been examined, with findings showing that spatial configuration and nonspatial factors, such as semantic expectations about destinations, significantly affect wayfinding performance [41]. Similarly, research by Li and Klippel [43] further emphasizes the impact of environmental legibility and familiarity on wayfinding behaviors in complex buildings. Their study showed that individual spatial abilities and familiarity with an environment significantly influenced the development of spatial knowledge, highlighting the importance of human and environmental factors in navigation. A cognitive–architectural description of circulation typology was introduced, employing graph-based methods for spatial analysis and comparing the layouts based on visibility and anticipated wayfinding difficulty [7]. The effects of different spatial metrics, such as entropy, the number of decision points, and segment integration, on the navigability of virtual environments were investigated [1]. All these factors contribute to the overall quantification of environmental complexity and inform this research on instruction validation.

In addition to the complexity of the overall layout, a separate stream of research has focused specifically on characterizing the complexity of individual settings within the environmental layout, including intersections. In the context of urban networks, the type and properties of urban intersections were used to quantify the complexity of a street network [20]. They considered *regular intersections* as those enabling turn choices following uniform turning angle increments across the $360°$ domain. Therefore, the angular distance is defined as the smallest angle needed to rotate the axes of an n-intersection to align with the edges of a perfect regular intersection (evenly spaced) (see Figure 1 and Equation 1)

$$\Delta = \Sigma_{i=1}^{n} \delta_i \quad (1)$$

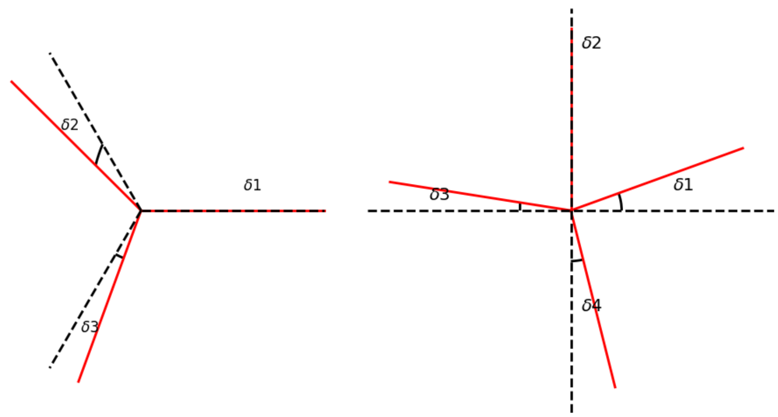


Figure 1: Red lines represent an actual indoor intersection, and dashed lines an imaginary perfect n-intersection. The angular distance of an intersection [20] is the minimum angle required to rotate its axis to become a perfect intersection (with the equally distributed axis around a full circle), therefore $\Delta = \Sigma_{i=1}^{n} \delta_i$. We rotate one axis to coincide with an axis of the imaginary perfect *n-intersection*. Here, (left) shows the aggregated angular difference for a 3-intersection $\Delta = \delta_1 + \delta_2 + \delta_3$, and (right) graph shows the aggregated angular difference for a 4-intersection $\Delta = \delta_1 + \delta_2 + \delta_3 + \delta_4$

2.4 Text-based games

Now that we have established floorplan complexity criteria using graph concepts, we need to express environment floorplans in a format that enables interaction—executability. The need to validate route instructions necessitates executing the directions without concerns for visual aspects in a controlled model. Therefore, we require a simple environment capable of interacting with text—interactive environments allow for dynamic and realistic execution of instructions. Converting static graphs or geometries into executable floorplans ensures that route instructions can be validated in an interactive setting. Text-based games are one form of executable environments that rely on symbolic modeling, capturing the interrelationships between spatial subdivisions or hierarchies of spatial entities.

Text-based games are environments where interactions, environment (world) definitions, and agent interactions are all conducted through text. Given that indoor space can be modelled using two approaches: using coordinates (geometric) and using abstract, descriptive concepts and their relationships (symbolic), text-based games are symbolic simulation environments of indoor space. Inform7 is a prominent text-based game authoring system, with environments modelled using predefined sets of hierarchically organised objects, grammar, rules, and location concepts such as rooms, directions, and regions [52]. Recently Microsoft Research introduced TextWorld [14], a sandbox environment for training reinforcement learning agents in text-based games, using an operational and logical framework based on Inform7 (shown in Figure S7 are examples of valid and invalid route instruction-following in TextWorld).

There are two main approaches to developing agents for text-based games: rule-based systems and reinforcement learning (RL) methods [34]. Rule-based systems use a set of explicit rules to dictate the actions of agents, while RL methods train the agent to take actions that maximize a reward signal. RL methods can be more effective for learning complex behaviors and adapting to new situations, but they can be more computationally intensive. Here in this study, we are using rule-based agents as the objective is to validate the route instructions deterministically.

For indoor navigation scenarios, geometric models of spaces work well for agents aware of their position, e.g. robots with sensory data. However, in text-based games agents interact purely through symbolic descriptions. The only existing method to systematically convert geometric floorplans into the format required by text-based game environments, complete with additional semantic spatial concepts and hierarchies required for modeling indoor spaces, is recently proposed [5].

3 Methodology

3.1 Universal validation model

The primary aim of this study is to develop a model that enables computationally determining the validity of route instructions in indoor environments. We propose a model with generic components that can be applied across different domains with the following characteristics:

- *Generic Validation Criteria*: The model assesses route instructions using four main criteria: understandability, executability, path-following, and destination guidance. These criteria apply universally across different environments and agent types. This is further discussed in Section 3.2;
- *Universal Simulation*: Any indoor floorplan that can be represented as an embedded graph can be fed into the validation model. The model ensures the representation of floorplans using a text formalism understandable as input in text-based games (Section 3.3);
- *Input Data Processing*: Floorplans and route instructions are the primary inputs of the framework. Floorplans are processed to extract decision points and path structures, while route instructions are parsed according to the specified grammar, elaborated in Section 4.3. To ensure consistent validation across various floorplans, the model normalizes the complexity of floorplans. Details are provided in Section 3.4

In our model, instructions are matched to choices in the environment that are deemed correct according to the formal specification of the grammar used. The model accommodates any turn-based grammar used in route instructions, each providing different levels of directional granularity. The model parses the route instructions according to the specified grammar (such as 4-sector, 6-sector, 8-sector, 16-sector, or CGG) and interprets each instruction within the grammar context to ensure that the agent can understand and execute the commands correctly. This ensures that the model is capable of incorporating additional grammar, such as a (theoretical) 32-sector or even asymmetric grammar, by simply updating the parsing (labels) and interpretation mapping.

This study employs a structured methodology comprising two main steps to evaluate the validity of route instructions for indoor navigation, shown in Figure 2.

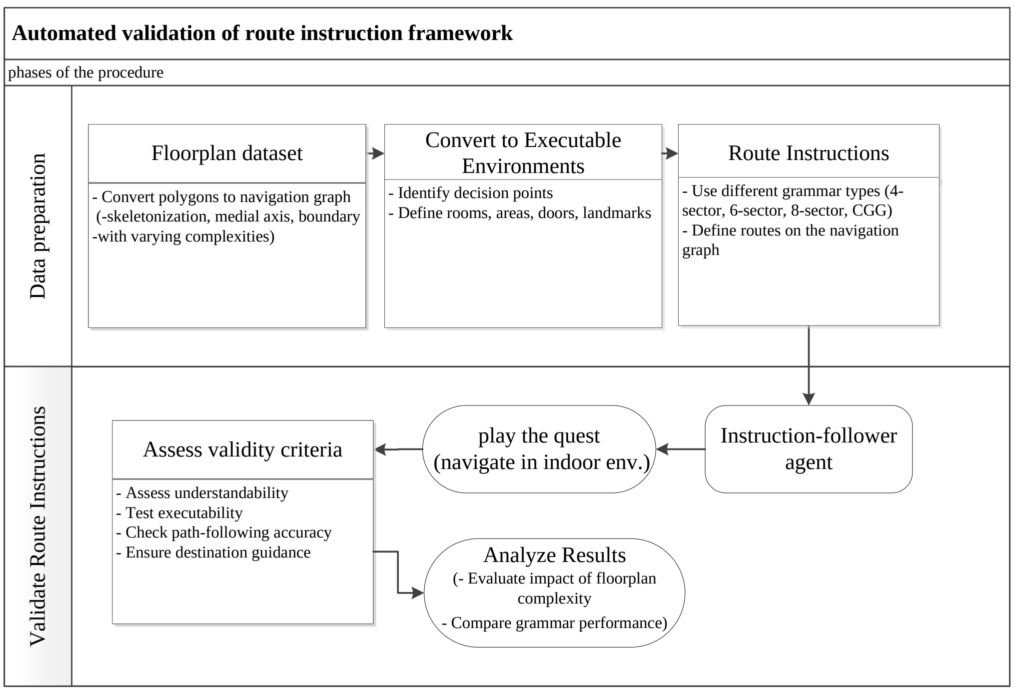


Figure 2: The workflow for automated validation of indoor route instructions across different floorplan and grammars (specific examples mentioned inside parenthesis are for demonstration only)

First, during *Data Preparation*, necessary data are prepared for validation. This includes converting floorplan datasets (e.g., polygons) into executable navigation environments. The conversion process involves the extraction of a navigation graph through e.g., skeletonization, medial axis computation, and/or boundary extraction (or any other alternative method). This results in navigation graphs representing environments with varying complexities. Additionally, decision points, rooms, areas, doors, and landmarks are identified in the environments and corresponded to the navigation graph. A grammar (including 4-sector, 6-sector, 8-sector, cognitively-grounded grammar (CGG) and 16 sectors) is then selected to express routes to be executed over the navigation graph.

In a second step, *Validation*, we use an instruction-following agent to navigate the text-based game environments. While the agent is capable of understanding verbal turn-based instructions (e.g., N/W/E/S and NE/NW/SE/SW, along with their egocentric equivalents), it is not aware of the specific intervals and angles used to generate the route instructions. For example, the agent does not know if 'north' corresponds to an interval of 80 to 100 or 45 to 135 degrees. The validity of the route instructions is assessed based on four criteria: *understandability*, *executability*, *path-based* route quality, and *destination guidance*. This comprehensive validation process allows for a detailed analysis of how floorplan complexity and grammar type impact navigation success, providing insights into the robustness and applicability of the proposed model. The main focus of this step involves the actual validation of route instructions, all examples noted below are just implementations for demonstration purposes.

3.2 Validation criteria

A universal model for validating route instructions across diverse indoor environments and instruction grammars requires well-defined metrics to evaluate the instructions. The following validation criteria consider the interplay between the agent's capabilities, the environment, and the linguistic formulation of the instructions. Establishing validity criteria supports developing a robust and generalizable validation model. We systematically assess route instructions based on the criteria below to ensure that the model provides reliable validation results:

- **Understandability**: *route instructions must be clear and comprehensible to the agents.* Thus, the grammars interpretable by agents must match those encoding the instructions. For example, if an agent can only understand egocentric directions, a route instruction *"go West"* would not be considered valid. We have developed agents that can comprehend egocentric (self-centered) and allocentric (environment-centered) directions, but here only report on experiments with egocentric directions—a direct mapping to allocentric directions would not alter the results.
- **Executability**: *route instructions must be executable by the agent, matching their capabilities and limitations.* Executability relates the agent's capabilities to the constraints imposed by environmental affordances [25, 53, 63]. For instance, an agent relying on a wheelchair is not able to follow instructions to climb stairs—such route instructions do not match the characteristics of the agent, one of the possible reasons for being not executable, and are therefore invalid. Text-based environments allow us to model these affordances and ensure the matches between agents and environments. In this study, agents were capable of executing all given instructions. Thus, the environments were controlled for executability, and they only constraints related to entering/exiting rooms via doors.
- **Path-based**: *route instructions must lead the agent along a defined* path *rather than a random* walk. Following their graph-theoretic meaning, following a *path* implies that the resulting instructions will not require the repeated traversal of an edge, and therefore the agent's route cannot be completed successfully purely by following e.g., a random walk of unconstrained length. Route instructions in this study describe instructions to follow the longest shortest path (Section 4.2).
- **Guidance to the Intended Destination**: *route instructions must guide the agent to the* intended *destination.* Only reaching the intended target is acceptable as a valid destination for the agent, i.e. the instructions are considered invalid even if a pragmatically equivalent surrogate is reached. Consider instructions to find *toilets* in a building—if the agent finds another cluster of restrooms then intended while navigating, this

would not constitute successfully reaching the destination. In this study, the agents continuously verify at every action (e.g., turn or decision point), whether they arrived at the instructed destination. If so, navigation ceases (even if additional instructions are present). We do not provide semantically identical destinations (confounders) in this study.

These four criteria are demonstrated in the Figure 3.

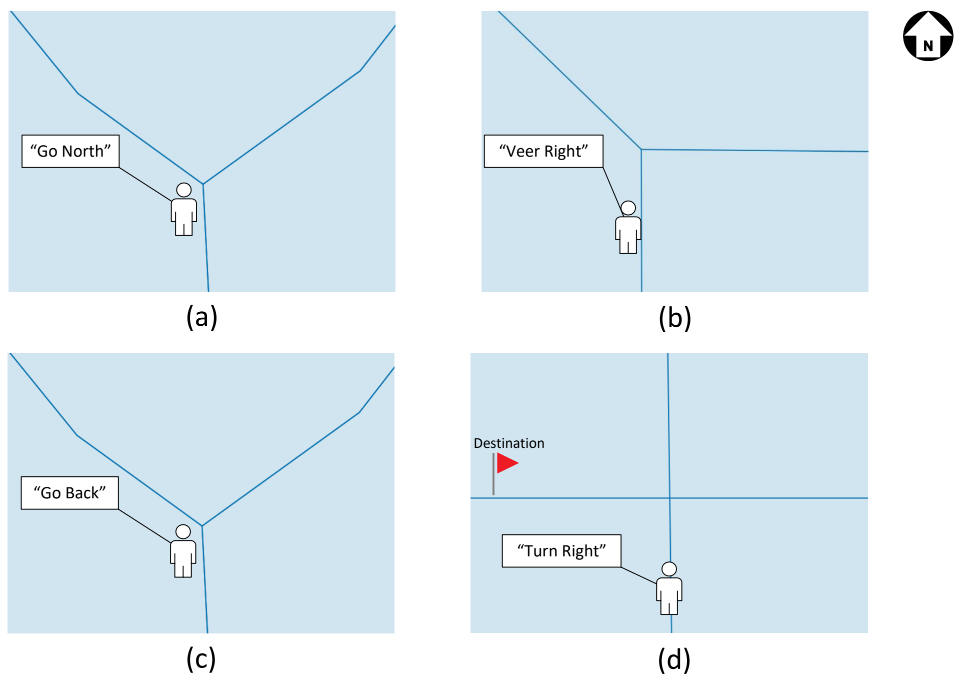


Figure 3: Examples of invalid route instructions based on different criteria, i.e., the agent cannot deterministically execute a route instruction to reach the intended destination (in all scenarios, it is assumed that the agent is facing north) : (a) *Understandability*: The agent is unable to comprehend allocentric cardinal directions (e.g., North, West, East, South) in an indoor environment, even though the route instruction provides them. (b) *Executability*: The agent understands *veer right* but cannot execute the turn due to a lack of a viable path. (c) *Path-Based*: *Going back* is clear and executable for the agent, but leads to a repetitive edge, i.e., the route is not a path. (d) *Guidance to Intended Destination*: The agent can grasp and follow the provided route instruction, but it won't lead to the desired destination.

In scenarios where route instructions contain ambiguities, such as multiple turns in the same direction within a single grammatical segment, our model identifies these conflicts and halts navigation. For example, an instruction like "go forward" in the case (a) of Figure 3 using a 4-sector grammar, creates uncertainty about which direction to follow. By detecting such ambiguities, the model prevents potential misnavigation and highlights areas where instructions may need refinement. Whenever ambiguous situations occur, we treat instructions as deterministically invalid (i.e., relying on a choice, and therefore resulting in a probabilistic outcome).

3.3 From floorplans to executable environments

To focus on the computational aspects of route instruction validation, we utilize a text-based environment (TextWorld). This approach deliberately excludes visual elements and treats all environmental features as symbolic, thereby removing the need to consider individual differences in visual perception and allowing for a controlled evaluation of instruction validity based on predefined textual criteria.

Following the only existing approach to systematically convert geometric floorplans into text-based game environments [5], we generate text-based games from the simulated floorplan environments. This approach allows for validating route instructions without requiring knowledge of coordinate positions. This method mirrors real-world indoor wayfinding where symbolic route instructions are provided to wayfinders localized by qualitative descriptions [3].

The text-based environment specifications include rooms, areas, doors, and landmarks. The process of generating specifications is divided into three steps:

1. Identifying decision points in navigation graphs;
2. Defining the rooms, areas, doors, and landmarks;
3. Identifying physical connections between areas (i.e. doors) or virtual connections between areas (i.e. Ulinks).

The specification of "indoor concepts" is enriched with the specification of rooms, areas, doors, and landmarks, all defined as follows [5,56]:

- `IndoorArea`: An atomic indoor space that can be entered and contain objects, and is part of an `IndoorRoom`, allowing agent interaction with contained objects;
- `IndoorRoom`: A container formed by aggregating `IndoorAreas`, bounded by real-world boundaries (e.g., walls), where movement inside doesn't require explicit actions, but movement between rooms does;
- `IndoorFloor`: A collection of horizontally collocated `IndoorRooms` forms an `IndoorFloor` (not directly affording object containment);
- `Door`: A link connecting adjacent `IndoorAreas` in different `IndoorRooms`, that can be walked-through and requiring explicit agent interaction to change from one `IndoorArea` and `IndoorRoom` to another;
- `Ulink`: A link connecting adjacent `IndoorAreas` within the same `IndoorRoom`, that can be walked-through and, when used, leading to a change of the agent's current `IndoorAreas`;
- `Landmark`: All objects located inside an `IndoorArea` are landmarks. They cannot be *entered* or *walked-through*.

3.4 Floorplan complexity

Finally, we require a means to quantify the complexity of indoor floorplans, in order to assess the hypothesis underpinning our validation framework. If it holds, an inverse correlation between floorplan complexity and the validity of route instructions for a given turn-based grammar should be identified. Let $G(V, E)$ be an undirected graph with vertices V and edges E where nodes represent decision points and edges represent possible paths between these points. The *angular distance (AD)* is a metric that evaluates the complexity of individual intersections [20]. This metric measures the difference between the actual bearing of directions at an intersection and the ideal bearings in a perfectly regular n-intersection if $n > 2$.

We quantify the overall complexity of the floorplans as the *aggregated angular distances (AAD)* (Figure 1). The aggregated angular distance is the sum of the angular distances of all intersections in the environment. Equation 2 defines the aggregated angular distances for an environment, $\Delta_{\text{floorplan}}$. In this equation, δ_{ij} represents the difference between the bearing of the ith direction of the jth intersection, denoted by θ_{ij}, and the bearing of the closest corresponding edge in an ideal n-intersection, denoted by θ^*_{ij}, where one of its edges aligns with one of the edges of the actual intersections.

$$\Delta_{\text{floorplan}} = \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij} \quad \text{where} \quad \delta_{ij} = \theta_{ij} - \theta^*_{ij} \tag{2}$$

While many metrics for evaluating complexity exist [such as ICD, intelligibility, angular distance, asymmetry], our choice of the Aggregated Angular Distance (AAD) metric is deliberate and well-suited to the objectives of our study. The AAD metric effectively captures the specific aspects of complexity of the environment that are directly relevant to turn complexity, an aspect required to assess the relationship between floorplan complexity and the validity of route instructions based on turn-based grammars.

4 Demonstration

4.1 Floorplans

We construct a dataset that encompasses the full spectrum of possible floorplan shape complexities with a particular emphasis on the complexities of indoor intersections, captured via the aggregated angular distance metric (see Section 3.4).

To obtain a dataset of high variety of layouts, we leverage the diversity of font glyphs (latin character shape outlines), which include polygonal geometries with and without holes, with convex and non-convex shapes, and the combinations of such shapes 4. Using font glyphs as environmental models is a well-established method in computational wayfinding research [3,50].

Such a synthetic dataset of glyph-based floorplans surpasses the complexity of real-world layouts, which is by design skewed toward low complexity—complex environments are unnecessarily challenging wayfinders. Thus, glyph-based floorplans are a means to obtain a systematically controlled dataset of floorplan complexities, to study the effects of layout complexity on wayfinding instruction understandability in function of instruction grammar. The findings then translate and inform our understanding of real-world floorplan navigability.

To intentionally introduce angular variety in the resulting graphs to cover the full spectrum of intersection complexities, three approaches to construct navigation graphs from glyph outlines are used: polygon skeletonization, medial axis extraction, and boundary outline extraction.

- *Polygon skeleton:* skeltonization is the result of thinning a polygon. Polygon skeletons of floorplans are graphs capturing the connectivity of indoor spaces [49, 58] The nodes in skeleton graphs represent navigation decision points [2]. We used the common Voronoi diagram decomposition [21] implemented in GRASS GIS [27] to create floorplan skeletons.
- *Polygon medial axis:* The medial axis $MA(P)$ of a polygon P is the set of all points within P that have at least two closest points on the boundary of P, and there is no other point on the boundary with a distance smaller than $d(p, q_i)$:

$$MA(P) = \{p \in P \mid \exists q_1, q_2 \in \partial P, q_1 \neq q_2, \text{ such that } d(p, q_1) = d(p, q_2)\} \tag{3}$$

  where p is a point within the polygon, and $d(p, q)$ denotes the Euclidean distance between points p and q.
- *Polygon outer boundary.* To generate right-angle navigation graphs based on our floorplan, we extract the outer boundary of sans-serif polygons. Here we consider vertices as decision points because we need at least two edges for an intersection and if the boundary is strictly composed of right angles, there may be fewer or no intersections within the boundary itself, which can lead to fewer turn instructions being necessary

Skeleton and medial axis methods produce graphs with varying floorplan complexity given AAD criteria, while outer boundaries of glyphs yield graphs with regular $90°$ geometries and right-angled turns at intersections, common in real-world building layouts.

4.2 Paths and instructions

For each navigation graph in the dataset we select an origin-destination (OD) pair to define the route for validation. For fair comparison across floorplans, the length of the route described by instructions should enable to capture the complexity of the floorplan. A route with a single decision point may still be simple even in a complex floorplan. We therefore compute the Longest-Shortest Path (LSP) between all origins and destinations in a floorplan. The length, here, is defined as the number of turns (i.e., equivalent to the number of instructions). The LSP helps us to single out a challenging route in a floorplan, increasing the likelihood of encountering complex intersections.

Recall navigation graph $G(V, E)$, and let $d(u, v)$ be the shortest topological distance between these vertices in the graph G. Then, the length of the longest shortest path in the graph is given by:

$$LSP = max\{d(u, v) | u, v \in V\} \tag{4}$$

We generate route instructions for the LSPs in the route graph (Figure 4). The route instruction generate include the initial orientation of the agent, the specification of the grammar tested (one of *4-sector, 6-sector, 8-sector, 16-sector* and *CGG* turn-based grammar) (Figure 2).

By default most text-based games (e.g., TextWorld [14]) lack the rich expressive capability to model environments with complex intersections (e.g., an intersection with five arms), and support only four cardinal directions. We have, therefore, modified the environment to support the rich 8-direction so that our floorplan models reflect the overall complexity of the real-world floorplans. Similarly, agents can also be designed to understand and execute different complexities of instructions (e.g., 4-sector, 6-sector, and 8-sector). As in text-based

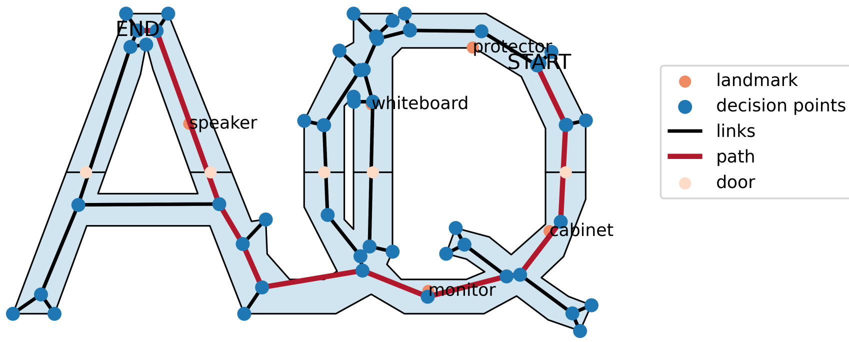


Figure 4: An example of a floorplan (glyphs of *"AJQ"*); decision points and connections are extracted using the medial axis.

games, agents do not perceive and categorize scenes, they must be *served* with scene specifications, i.e., they rely on route instructions and scene descriptions. As shown in Figure 5, here agents were enabled to understand and execute all eight cardinal and inter-cardinal directions.

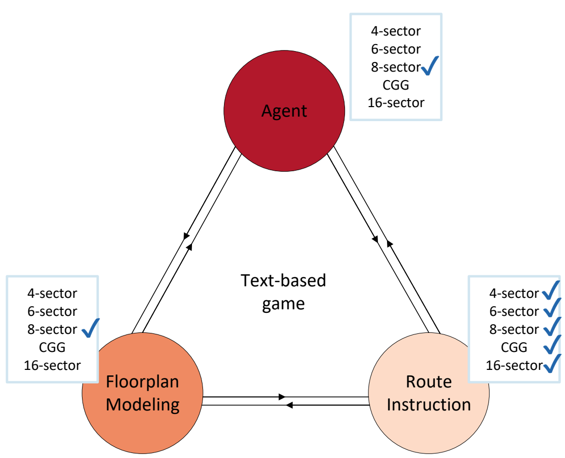


Figure 5: Relationship of environment modeling approaches, route instruction generation and agent's route following abilities based on grammars. Check-marks note elements implemented in this paper.

4.3 Grammar types

With our floorplans now transformed into executable environments, we can effectively test and validate route instructions with different grammar. Incorporating various grammars into our universal validation framework ensures that the model can accommodate and evaluate instructions across a broad spectrum of contexts. Grammar types are necessary to capture the varying levels of granularity and specificity needed for navigation scenarios with different correspondence between the language terms and turn angles.

The framework enables the validation of route instructions using any turn-based grammar type, showcasing this capability with four specific examples: 4-sector, 6-sector, 8-sector, CGG, and 16-sector grammar. Each grammar type offers a different level of granularity in describing directions. The 4-sector grammar divides orientations into four broad categories (front, left, right, and back), while the 6-sector grammar provides a more refined breakdown by adding intermediate directions. The 8-sector grammar further increases specificity with eight distinct orientations, and CGG, grounded in cognitive observations, offers nuanced instructions based on turn angles.

While we may use terms like "go slightly right" for the 8-sector grammar, we do not have generally accepted linguistic terms for the intermediate directions needed to afford the 16-sector grammar, which presents a barrier in indoor settings due to its complexity. Nevertheless, we have tested the 16-sector grammar using arbitrary linguistic terms such as "go very slightly right" to evaluate whether any improvement in validation occurs. The specifications for these grammar types, including their associated directional angles, are detailed in Table 2 and demonstrated in 6.

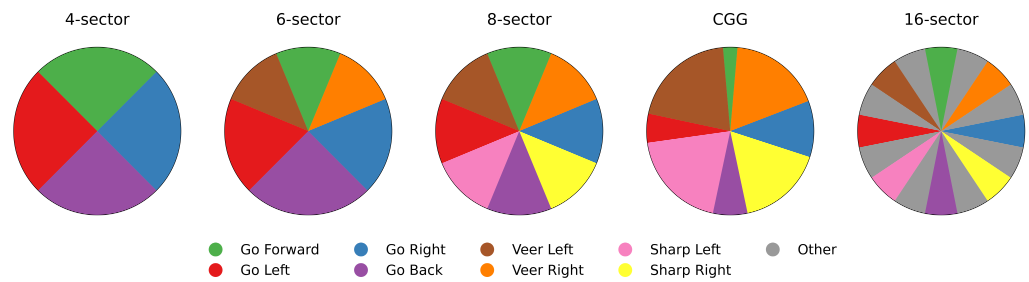


Figure 6: Route instructions and the corresponding turn angles in different grammar.

The CGG uses the same turn designations as the 8-sector grammar but deviates in the turn angles associated. From the agent's perspective, the labels of the CGG are nominally equivalent to those of the 8-sector grammar.

4.4 Dataset overview

A dataset of simulated floorplans was created using single or combinations of multiple glyphs (two and three characters), using serif, sans-serif, and right-angle rectilinear fonts (Table 3). The minimum, mean, and maximum number of rooms in a floorplan in our dataset are 2, 3.7, and 12, respectively. The resulting polygons are stored as different

Directions	4-Sector	6-Sector [69]	8-Sector	CGG [38]	16-Sector [59]
North	315-45	337.5-22.5	337.5-22.5	355-5	348.75-11.25
North-North-East	-	-	-	-	11.25-33.75
North-East	-	22.5-67.5	22.5-67.5	5-78	33.75-56.25
East-North-East	-	-	-	-	56.25-78.75
East	225-315	67.5-157.5	67.5-112.5	78-98	78.75-101.25
East-South-East	-	-	-	-	101.25-123.75
South-East	-	-	112.5-157.5	98-168	123.75-146.25
South-South-East	-	-	-	-	146.25-168.75
South	135-225	157.5-202.5	157.5-202.5	168-192	168.75-191.25
South-South-West	-	-	-	-	191.25-213.75
South-West	-	-	202.5-247.5	192-252	213.75-236.25
West-South-West	-	-	-	-	236.25-258.75
West	45-135	202.5-292.5	247.5-292.5	252-291	258.75-281.25
West-North-West	-	-	-	-	281.25-303.75
North-West	-	292.5-337.5	292.5-337.5	291-355	303.75-326.25
North-North-West	-	-	-	-	326.25-348.75

Table 2: Associated turn angle intervals (in degrees) for different grammars, measured clockwise from straight (0°). In our implementation, we have converted the instructions to *egocentric directions* (e.g., straight, turn slight right, turn very sharp left). However, our text-based environment is capable of understanding egocentric and cardinal (e.g., North, East, South, West) as well as intercardinal directions (e.g., Northeast, Northwest).

.

`IndoorRoom`s, together forming a floor in a single-story building (conceptually, a single `IndoorFloor`).

Parameter	Value
Fonts	3
Graph models	3
Route instructions' grammar	4
One-glyph floorplans	6
Two-glyph floorplans	70
Three-glyph floorplans	115
Total unique floorplan layouts	191
Total floorplans	742
Total route instructions	2968

Table 3: Description of the virtual floorplan dataset.

Floorplans with skeleton axis graphs have the highest LSP lengths due to their inherent complexity, while the presence of 90-degree angles along the boundaries of rectilinear environments leads to route instructions with right-angle turns, covered by the basic turn concepts common to all tested grammars (Figure 7).

4.5 Implementation

The *aggregated angular distance* complexity for each floorplan is normalized into a scale of 0 to 1, using the normalization equation of $Y_{normalized} = (Y - Y_{min})/(Y_{max} - Y_{min})$, to enable meaningful comparisons across the dataset. Floorplans are categorized by angular distance into bins with increments of 0.1 (Figure 7, with detailed histograms in Figure 8). Right-angle floorplans cluster at the lower end of the aggregated angular distance complexity spectrum due to the consistent 0 complexity of routes with only regular, 90-degree turns when navigation graphs are constructed from boundaries and with highly regular n-intersections when constructed from medial axes.

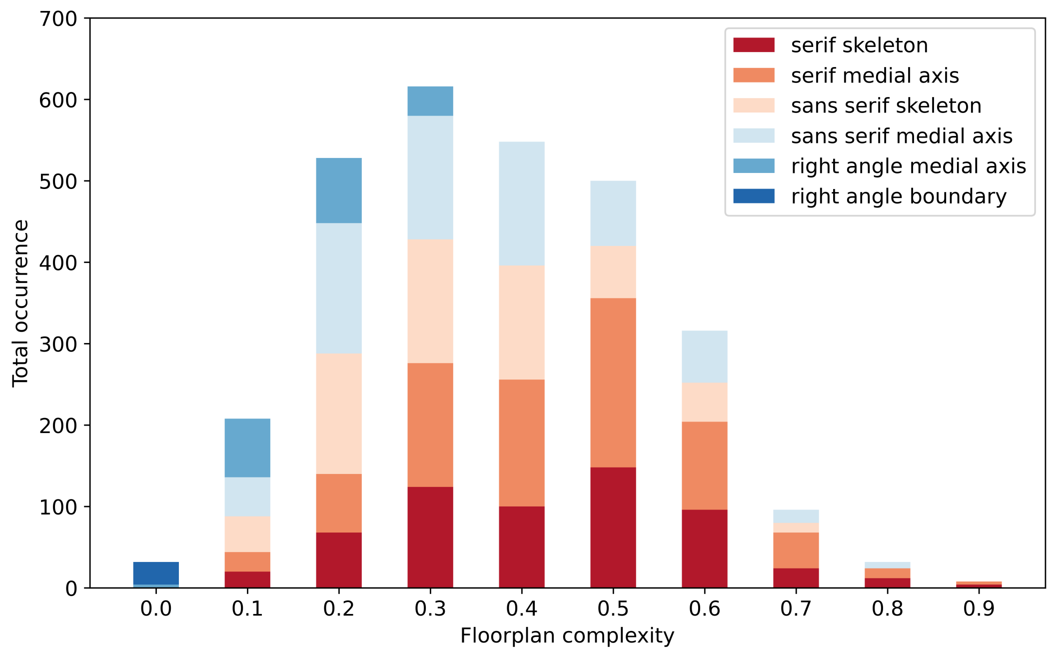


Figure 7: Histogram of floorplan complexity in the dataset

In contrast, none of the environments within the *serif* and *sans serif* categories exhibit zero overall complexity (the lowest measured complexity is 0.1). Notably, the histograms of *serif* environments yields higher average complexities (blue lines in Figure 8) compared to the *sans serif* environments and right-angle environments. This is because the navigation graphs inherit geometric features of the subtle serifs of the glyphs.

We generate a dataset of route instructions for validation. For every floorplan, the route following the longest shortest path (LSP) is calculated using the *networkx* Python library [29], resulting in instructions for each letter(s) combinations, origin, destination, and grammar combination (an entry of the data is shown in Figure 9).

We then generate inputs of Inform7-compatible specification executable in the TextWorld environment [14]. These inputs contain the floorplan geometry, the route origin, and the agent's primary orientation. With the room, door, landmark, and skeleton data corresponding to the specified floorplan, we create objects representing `IndoorRoom`, `IndoorArea`, `Door`, `Landmark`, and `Ulink` (links between areas). To validate route instructions, we execute the generated route instructions step by step, receiving feedback from the game environment upon each instruction.

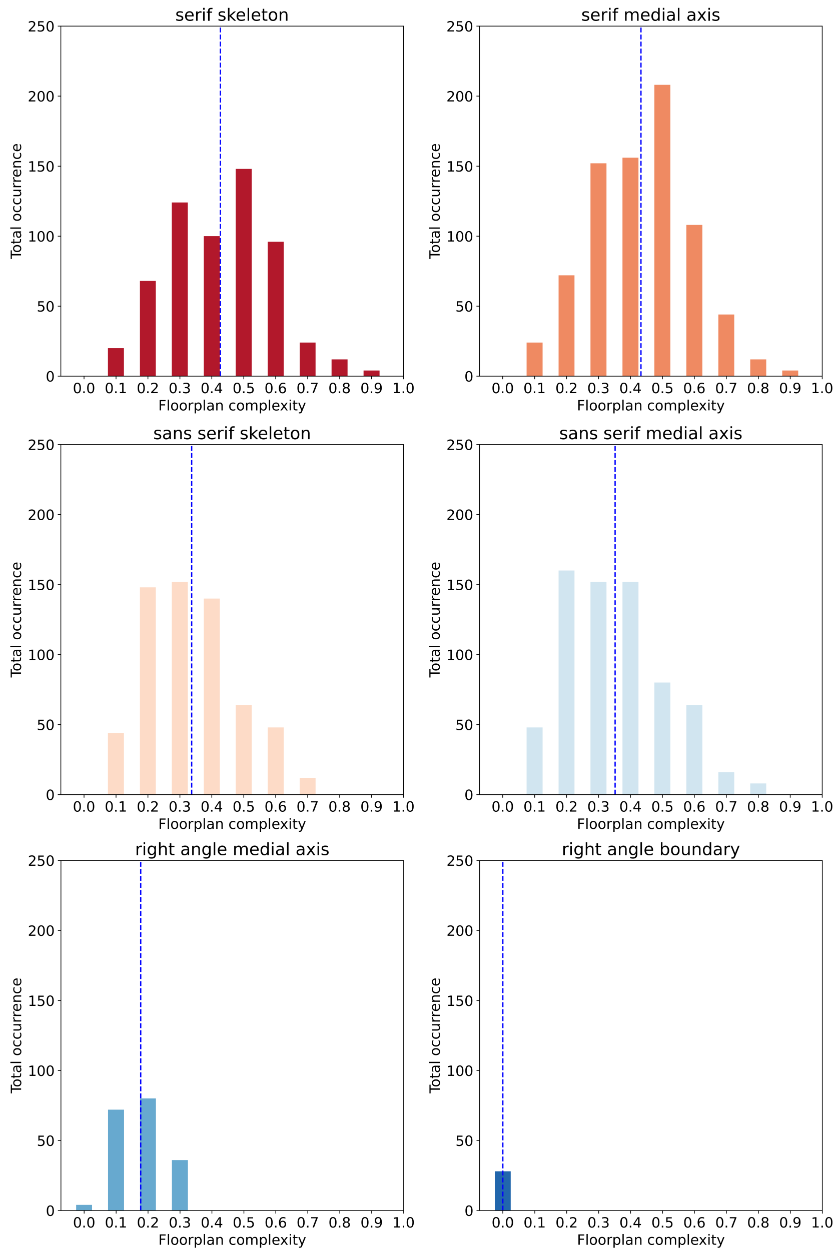


Figure 8: Categorized histograms of each combination of fonts and graph extraction vs. their frequency of occurrence on each interval of floorplan complexity. Blue lines signify the x value of the average.

```xml
<letter name="AEJ_Average-Regular">
    <route x_origin="502701.0" y_origin="1007053.0" x_destination="500009.0" y_destination="1007591.0">
        <style name="Turn-based">
            <grammar name="4sector">
                <route_instruction>Turn left. Go straight. Go straight. Turn left. Turn left. Go straight. Turn left. Turn right. Turn left.
                 Turn left. Turn right. Turn left. Go straight. Arrive at destination.</route_instruction>
            </grammar>
            <grammar name="6sector">
                <route_instruction>Turn left. Turn slightly right. Turn slightly right. Turn slightly left. Turn left. Turn slightly right.
                Turn slightly left. Turn right. Turn left. Turn left. Turn right.
                Turn slightly left. Go straight. Arrive at destination.</route_instruction>
            </grammar>
            <grammar name="8sector">
                <route_instruction>Turn left. Turn slightly right. Turn slightly right. Turn slightly left. Make a sharp left.
                 Turn slightly right. Turn slightly left. Turn right. Turn left. Turn left. Turn right. Turn slightly left.
                 Go straight. Arrive at destination.</route_instruction>
            </grammar>
            <grammar name="Klippel">
                <route_instruction>Make a sharp left. Turn slightly right. Turn slightly right. Turn slightly left. Make a sharp left.
                Turn slightly right. Turn slightly left. Turn right. Turn left. Turn slightly left. Turn right. Turn slightly left.
                 Turn slightly right. Arrive at destination.</route_instruction>
            </grammar>
        </style>
```


Figure 9: An entry of the generated route instructions for the floorplan dataset

We monitor whether the agent has arrived at its destination by comparing the game state feedback with the destination specified in the instructions. If the agent successfully reaches the destination, the instruction is considered valid; otherwise, it is marked invalid (see invalid scenarios in Figure 3). No navigation heuristics have been incorporated, resulting in purely deterministic validation of instructions.

4.6 Demonstration on real-world floorplans

To validate the applicability of our model to real-world environments, we selected a diverse set of floorplans from prominent shopping malls worldwide used by [6]. These floorplans were manually digitized and processed for analysis using our approach. The details of each selected floorplan, including their size and architectural complexity, are summarized in Table 4. To ensure variety in our evaluation, we selected malls that represented a wide range of design complexities and geographical regions. This allowed us to analyze environments with different design complexities. The floorplans were digitized using QGIS [1] and converted into navigation graphs through the use of *NetworkX* [29] and the *momepy* python library [19]. These graphs served as inputs for our validation model.

Our analysis reaffirms that real-world floorplans tend to exhibit lower aggregated angular distance complexity when compared to the synthesized floorplans generated for this study. This supports the rationale for creating synthetic datasets to explore a wider range of complexities beyond what is typically encountered in real-world environments. The results demonstrate the versatility of our approach in handling real-world and highly complex synthesized floorplans, effectively supporting our model's broad applicability.

5 Results and discussion

The results presented in Figure 11 depict the validity of route instructions for LSP routes within floorplans of varying complexity.

[1] www.qgis.org

Table 4: Summary of real-world shopping mall floorplans [6]

Name	City	Number of Shops	Aggregated Angular Distance
Chadstone	Melbourne (Australia)	530	0.2
Emporia	Malmo (Sweden)	220	0.1
International Village Mall	Vancouver (Canada)	76	0.1
Falcon Mall (under construction)	Dubai (UAE)	61	0.1

The results in Figure 11 indicate that the *8-sector* grammar enables agents to navigate with the highest success rate, i.e., it is the grammar resulting in the highest occurence of valid route instructions. The *6-sector* grammar follows, showcasing the grammar's ability to provide successful navigation guidance. This is because in most environments, and therefore routes, the agent rarely needs to retrace (use the *go back* instruction) or be exposed to the *go sharp left*, and *go sharp right* directional instructions. Yet, these are the significant additions in the 8-sector with respect to the 6-sector grammars. On the other hand, the *CGG* demonstrates a moderate success rate. This could be due to the asymmetric distribution of directions with more focus on inter-cardinal directions. Lastly, the *4-sector* grammar exhibits the lowest percentage of valid instructions, failing to adequately describe non-right-angle turns.

The main hypothesis of the study is that there is a negative correlation between overall floorplan complexity and route instruction validity, with the relationship modulated by the grammar used. We evaluate this hypothesis using a regression model between the floorplan complexity and the validity of route instructions for different grammars, and interpret the results from Table 5 and Figure 11 as follows:

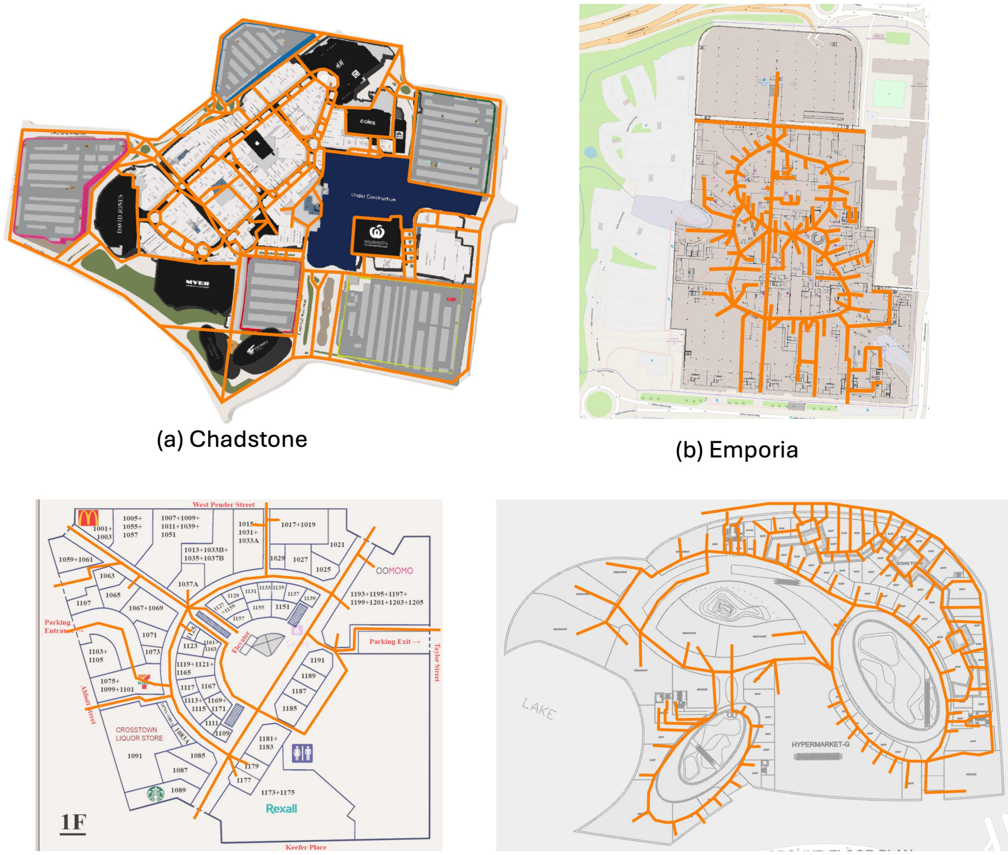
The formulation of the alternative hypotheses to the principal hypothesis is:

- *Null Hypothesis (H_0)*: There is no correlation (no linear relationship) between the overall floorplan complexity and validity of route instructions; vs.
- *Alternative Hypothesis (H_1)*: There is a non-zero linear relationship between the overall floorplan complexity and validity of route instructions

F-statistic is used to test the overall significance of the regression model. DOF_1 and DOF_2 are degrees of freedom for the regression and residuals, respectively. Here, $DOF_1 = 1$ because there is only one predictor variable (overall complexity level) in our model (the slope is the parameter to be estimated). DOF_2 is equal to the total number of data points minus the number of parameters being estimated (including the intercept term). In our case of n data points, one predictor variable, and one intercept, $DOF_2 = n - 2$. We chose the significance level (α) of 0.05. Given the DOF_1, DOF_2, and (α), the critical F-value is approximately 5.32. If the calculated F-value is greater than the critical F-value, the null hypothesis is rejected, i.e., the linear relationship is statistically significant.

Based on these results, we summarize our findings:

- **4-sector:** The performance of the 4-sector grammar for navigation in indoor environments is high if the environment is simple (i.e. only contains right angle turns).



(a) Chadstone

(b) Emporia

(c) International Village

(d) Falcon Mall

Figure 10: Examples of real-world shopping mall floorplans: (a) Chadstone Mall in Melbourne, Australia, with 530 retail units and an aggregated angular distance of 0.2; (b) Emporia Mall in Malmo, Sweden, featuring 220 stores and an aggregated angular distance of 0.1; (c) International Village Mall in Vancouver, Canada, containing 76 shops and an aggregated angular distance of 0.1; (d) Falcon Mall in Dubai, UAE, currently under construction, with 61 stores and an aggregated angular distance of 0.1. The red lines illustrate the digitized navigation paths. (Credits of basemaps belongs to a: www.chadstone.com.au, b: www.emporia.steenstrom.se, c: www.internationalvillagemall.ca, and d: www.falconcity.com)

While the 4-sector grammar may be adequate for instructions in highly regular environments, it is rarely adequate and reliable for more complex environments. As indicated by the F-value of 3.15, the correlation coefficient suggests a weak linear relationship between overall floorplan complexity levels and the validity of route instruction generated by the 4-sector grammar. The negative slope of -55.3 further indicates that as the overall floorplan complexity level increases, the fraction of valid instruction tends to decrease. However, the F-value of 3.15 for statistical significance

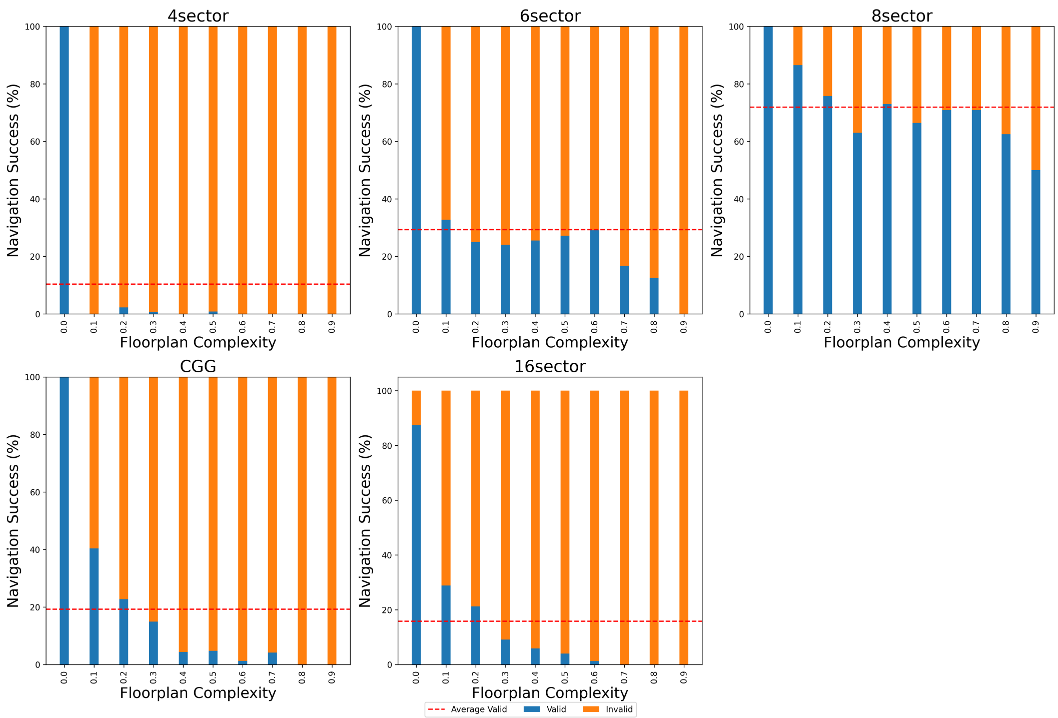


Figure 11: Validity of route instructions based on the grammar in floorplans categorized by their Δ complexity for LSP (Longest Shortest Path); the percentage of valid route instructions are 54.2% for 8-sector, 31.3% for 6-sector, 11.4% for CGG, and 3.1% for 4-sector grammars.

Grammar	F-value	Slope
4-sector	3.15	-55.3
6-sector	9.40	-64.6
8-sector	18.50	-37.9
CGG	12.09	-79.7
16-sector	10.96	-67.9

Table 5: Statistical significance of regression model of overall floorplan complexity measure vs validity of route instructions

suggests it may not be as statistically robust as the relationships observed with other grammars.

- **6-sector:** The 6-sector grammar exhibits a more pronounced relationship between overall floorplan complexity levels and valid percentages compared to the 4-sector grammar. With an F-value of 9.40, the correlation is stronger, indicating that the 6-sector grammar is more sensitive to changes in environmental overall floorplan complexity. The negative slope of -64.6 suggests a substantial decrease in valid percentages with increasing overall floorplan complexity. This indicates that the 6-sector grammar may struggle to provide reliable instructions in more intricate environments.
- **8-sector:** In contrast to the 4-sector and 6-sector grammars, the 8-sector grammar displays a significantly stronger correlation between overall floorplan complexity levels and valid percentages, as evidenced by the high F-value of 18.50. The negative slope of -37.9 indicates a clear negative trend (although with the least deterioration of all tested grammars), suggesting a decrease in the validity of instructions with increasing overall floorplan complexity. This may imply that the 8-sector grammar is more adaptable to moderately complex environments but still faces challenges in highly intricate settings.
- **CGG:** The CGG demonstrates a slope of -79.7, indicating a sharp decrease in valid percentages as overall floorplan complexity levels rise, suggesting that the CGG is particularly sensitive to variations in environmental overall floorplan complexity (F-value of 12.09). This grammar may excel in providing accurate instructions in simple, structured environments but struggles in complex settings with irregular turn angles. This may be because of the irregular angular discrimination of turns, with heavily nuanced discrimination towards a narrow set of angles in front of the wayfinder.
- **16-sector:** The 16-sector grammar demonstrates a considerable sensitivity to variations in overall floorplan complexity, as reflected by an F-value of 10.96. The negative slope of -67.9 suggests that valid percentages decrease substantially as the complexity of the floorplan increases. This indicates that while the 16-sector grammar is capable of providing detailed and nuanced instructions in less complex environments, it struggles to maintain reliability in highly intricate settings. The higher resolution of directional instructions inherent to the 16-sector grammar may introduce additional complexity, making it less effective in complex environments where simpler, more general instructions might be more beneficial. One potential reason for the increased number of failures with the 16-sector grammar, compared to the 8-sector grammar, could be the agent's limited capability to differentiate between very slight variations in directions (e.g., turn very slight right versus turn slight right) when con-

veyed through verbal instructions. This granularity may exceed the agent's ability to interpret subtle angular distinctions, leading to confusion and decreased navigation accuracy.

Overall, these findings highlight the importance of considering the choice of grammar in indoor navigation systems, particularly in environments with varying levels of overall floorplan complexity. While certain grammars may perform well in straightforward and simple settings, their efficacy may diminish in more complex environments.

The percentages of valid route instructions and the average success rate (red dashed line in Figure 11) reveal insights into the performance of different grammars in indoor wayfinding scenarios.

Firstly, the 4-sector grammar exhibits relatively stable performance across all levels of floorplan complexity but struggles significantly in more complex environments. The percentage of valid instructions remains low, indicating that this simpler grammar is insufficient for capturing the nuances of more intricate floorplans. The 6-sector grammar, while showing an improvement over the 4-sector, also sees a notable drop in the validity of instructions as floorplan complexity increases, reflecting its limitations in highly detailed settings. This trend suggests that while adding more sectors can improve performance, this is not sufficient for highly complex environments.

The 8-sector grammar shows a marked improvement in handling complex environments compared to the 4-sector and 6-sector grammars. This grammar demonstrates a higher resolution of directional instructions, which translates into higher validity in more complex settings. The correlation between increased floorplan complexity and the validity of instructions is stronger, indicating that the 8-sector grammar is better suited for intricate environments. This supports the notion that a more nuanced grammar can indeed lead to better performance in complex scenarios, compared to 4-sector and 6-sector grammar.

In our analysis of various grammatical structures for route instructions, we observed an intriguing pattern. The 16-sector and CGG were designed to provide more detailed directional instructions, theoretically offering greater precision in navigation. However, our results reveal an unexpected trend: compared to the 8-sector grammar, these more complex systems actually show a decline in the generation of valid route instructions as floorplan complexity increases. This finding suggests that *there is an optimal level of grammatical complexity for generating valid route instructions*. The 8-sector grammar, with its balance of detail and simplicity, appears to hit this sweet spot, outperforming simpler and more complex alternatives across various levels of environmental complexity. This optimal performance is evident in its higher F-value (18.50) and less steep decline in effectiveness (-37.9 slope) compared to other grammars. These results indicate that while increased grammatical complexity can provide more nuanced directional information, it may also introduce challenges in practical application, particularly in more complex environments. Thus, our study points to the importance of finding a balance between descriptive power and practical usability in the design of navigational grammars

These findings reinforce the anticipated but previously unconfirmed benefits of nuanced instruction grammars for complex indoor layouts and the robust applicability of simple instruction grammars in environments constrained to right angles. While it may seem that we have confirmed an obvious relationship between grammar complexity and floorplan complexity, this consistency in results actually underscores the reliability and effectiveness of our validation framework. The somewhat predictable outcomes indicate that the validation model performs as expected when applied to known scenarios, validating its robustness. By successfully demonstrating the correlation between floorplan complexity and the richness of the grammar, we confirm that our framework can reliably assess route instructions. This reinforces our main contribution: the development of a universal, automated validation model for route instructions that consistently performs across various environments and grammatical complexities.

Floorplan	Nodes	LSP Length (m)	Edges in LSP	4S	6S	8S	16S	CGG
Falcon	151	299.7	11	0	0	1	0	1
Chadstone	259	1019.8	20	0	0	1	0	0
IVM	45	181.7	24	0	1	1	1	1
Emporia	67	530.7	23	0	0	1	1	1

Table 6: Summary of Floorplan Data with Grammar Evaluations. 4S represents 4-sectors grammar, 6S represents 6-sectors, 8S represents 8-sectors, 16S represents 16-sectors, and CGG represents cognitively grounded grammar. Values indicate the validity (1) or invalidity (0) of the corresponding grammar.

Our dataset consists of both real-world and synthetic floorplans, specifically designed to encompass and exceed the levels of complexity encountered in real-world environments. Although the selected shopping malls are quite large, the complexities of their floorplans tend to cluster towards the lower end of the spectrum, with aggregated angular distances typically ranging between 0.1 and 0.2. In contrast, our synthetic dataset spans a broader range of complexities, extending up to 1, thus enabling a more comprehensive analysis. By addressing the validation problem under the most challenging conditions–where synthetic floorplans represent the highest complexities–we demonstrate that our approach is robust and capable of handling a variety of potential scenarios.

To evaluate the applicability of our validation framework beyond synthetic environments, we applied our grammatical structures to a set of real-world floorplans, as summarized in Table 6.

The results indicate that the 8-sector grammar consistently outperforms both simpler and more complex grammars across diverse real-world settings. For instance, in the Falcon and Emporia floorplans, the 8-sector grammar successfully generated valid route instructions, whereas the 4-sector and 6-sector grammars failed to do so. Similarly, the CGG showed limited success, matching the performance of the 8-sector grammar in some cases but falling short in others, such as Chadstone. Notably, the 16-sector grammar did not yield any valid instructions in half of the floorplans, aligning with our synthetic data findings that increased grammatical complexity does not necessarily translate to improved performance in real-world scenarios. The IVM floorplan presented a unique case where the 6-sector and CGG were able to generate valid instructions, suggesting that certain real-world layouts may benefit from specific grammatical structures. Overall, these real-world implementations corroborate our synthetic data results, reinforcing the conclusion that the 8-sector grammar strikes an optimal balance between detail and usability. This consistency across synthetic and real-world data underscores the robustness and generalizability of our validation framework, demonstrating its effectiveness in diverse indoor environments.

6 Conclusion and future work

This study computationally determined how the relationship between indoor floorplan complexity, route instruction grammar, and the validity of route instructions in indoor environments can be understood. 8-sector and 6-sector showed the most valid grammar, while more or less, 4-sector, 6-sector, and CGG exhibited a negative correlation between overall floorplan complexity and route instruction validity. Our work makes two main contributions, which we discuss in detail below:

- *Proposal of a universal validation model:* We proposed a computational model that can be applied universally to a diverse set of environments and turn-based grammar for assessing route instruction validity. This model offers a robust framework, moving away from subjective human validation to a reliable computational approach.
- *Systematization of route instruction validity criteria:* By leveraging the systematization of requirements for the validity of route instructions, The framework's ability to produce consistent and predictable results under stress tests underscores its reliability and generalizability. These understandability, executability, path-following accuracy, and destination guidance criteria ensure that route instructions are linguistically sound and pragmatically effective for real-world navigation.
- *Development and application of a comprehensive floorplan dataset* Additionally, we developed and applied a new synthetic dataset of indoor floorplans, specifically designed to cover and exceed the spectrum of complexities found in real-world environments. This dataset allowed us to systematically assess the impact of floorplan complexity on the validity of route instructions. By incorporating various geometric transformations and using diverse font glyphs, we created a wide range of intersection complexities. The results from our dataset confirmed the anticipated correlation between grammar complexity and floorplan complexity. However, our findings also revealed a breaking point where further increasing the grammatical complexity no longer improves navigation success rates and may even reduce validity in highly complex environments. Our findings validate the initial hypothesis, and the results further confirm the effectiveness of our validation framework.

In conclusion, this systematic approach not only demonstrated the robustness of our model but also provided insights into the relationship between environment complexity and route instruction validity. This research contributes to understanding the relationship between grammar, overall floorplan complexity, and the validity of route instructions in indoor navigation systems. These findings reinforce the anticipated but previously unconfirmed benefits of nuanced instruction grammars for complex indoor layouts and the robust applicability of simple instruction grammars in environments constrained to right angles. The consistency in results underscores the effectiveness of our validation framework, serving as evidence that our proposed model works consistently under stress tests. By successfully demonstrating the correlation between floorplan complexity and the richness of the grammar, we confirm that our framework can reliably assess route instructions across any indoor environment expressible as a navigation graph and any route instruction that uses turn-based grammar.

Future research should investigate the assessment of route instruction validity with agents capable of heuristics, e.g., in instances of uncertainty at decision points. This could involve the agent selecting the path closest to the instructed direction based on turn angle. Additionally, future work could incorporate individual differences and previous navigation knowledge into the validation process, moving beyond deterministic execution of instructions. Agents capable of learning and adapting based on experience may provide more realistic models of human wayfinding. Similarly, future research could incorporate more aspects of route instructions, such as linguistic structures and contextual information. Landmarks complement turn instructions to reduce ambiguity, and the study of landmark-based instructions is subject to future work. Another future research involves exploring the correlation between the validity of route instructions and floorplan complexity, considering scenarios where agents may not possess full capabilities, in terms of grammar understanding and execution, as well as in the modeling of the environment. An extension of the approach to multi-storey buildings will be considered in future work. We focus on deterministic validity derived from explicitly outlined actions in route instructions. Consequently, partial and probabilistic validity are beyond the scope of this study and are considered potential areas for future research. Future work could also extend the current model by incorporating these perceptual elements, such as the ability to distinguish between similar objects in the environment (e.g., identifying a 'green door' among other colored doors) and handling signage as part of route instructions.

While the current study focuses on the binary assessment of route instruction validity, future work could explore the relationship between instruction complexity and human navigation success. Investigating whether there exists a correlation between the validity of instructions and navigation outcomes would enhance the applicability of our method. Additionally, developing a probabilistic assessment framework could provide a more detailed understanding of instruction complexity and its impact on user performance. Moreover, while our method provides a robust computational validation of route instructions based on criteria such as understandability, executability, path-based correctness, and guidance to the intended destination, it does not currently incorporate human evaluations. Future research could extend this work by conducting user studies to assess the practical effectiveness and user satisfaction of the validated instructions. Such studies would complement our computational approach and provide a more comprehensive evaluation of route instruction quality.

7 Data availability statement

All data and codes are available in the repository at the following link https://doi.org/10.6084/m9.figshare.24208518.

References

[1] Entropy and a sub-group of geometric measures of paths predict the navigability of an environment. *Cognition 236* (2023), 105443. doi:10.1016/j.cognition.2023.105443.

[2] AMOOZANDEH, K., WINTER, S., AND TOMKO, M. Space decomposition based on visible objects in an indoor environment. *Environment and Planning B: Urban Analytics and City Science 49*, 3 (2022), 883–897. doi:10.1177/23998083211037347.

[3] AMOOZANDEH, K., WINTER, S., AND TOMKO, M. Granularity of origins and clustering destinations in indoor wayfinding. *Computers, Environment and Urban Systems 99* (2023), 101891. doi:10.1016/j.compenvurbsys.2022.101891.

[4] AMORES, D., TANIN, E., AND VASARDANI, M. A proactive route planning approach to navigation errors. *International Journal of Geographical Information Science 35*, 6 (2021), 1094–1130. doi:10.1080/13658816.2020.1820508.

[5] ARABSHEIBANI, R., HAMZEI, E., AMOOZANDEH, K., WINTER, S., AND TOMKO, M. From floorplan to navigation concepts: Automatic generation of text-based games. *AGILE: GIScience Series 4* (2023), 2. doi:10.5194/agile-giss-4-2-2023.

[6] ARABSHEIBANI, R., HAUNERT, J.-H., WINTER, S., AND TOMKO, M. Strategic allocation of landmarks to reduce uncertainty in indoor navigation. *Computers, Environment and Urban Systems 114* (2024), 102198. doi:10.1016/j.compenvurbsys.2024.102198.

[7] ASYA NATAPOV, SASKIA KULIGA, R. C. D., AND HÖLSCHER, C. Linking building-circulation typology and wayfinding: design, spatial analysis, and anticipated wayfinding difficulty of circulation types. *Architectural Science Review 63*, 1 (2020), 34–46. doi:10.1080/00038628.2019.1675041.

[8] BEST, G. Direction-finding in large buildings. architectural psychology. *London: RIBA Publications, S 72* (1970), 75.

[9] BLUKIS, V., PAXTON, C., FOX, D., GARG, A., AND ARTZI, Y. A persistent spatial semantic representation for high-level natural language instruction execution. In *Proceedings of the 5th Conference on Robot Learning* (08–11 Nov 2022), A. Faust, D. Hsu, and G. Neumann, Eds., vol. 164 of *Proceedings of Machine Learning Research*, PMLR, pp. 706–717.

[10] BRUNYÉ, T. T., GAGNON, S. A., GARDONY, A. L., GOPAL, N., HOLMES, A., TAYLOR, H. A., AND TENBRINK, T. Where did it come from, where do you go? direction sources influence navigation decisions during spatial uncertainty. *Quarterly Journal of Experimental Psychology 68*, 3 (2015), 585–607. doi:10.1080/17470218.2014.963131. PMID: 25285995.

[11] BUGMANN, G., KLEIN, E., LAURIA, S., AND KYRIACOU, T. Corpus-based robotics: A route instruction example. In *Proceedings of intelligent autonomous systems* (2004), pp. 96–103.

[12] CHEN, H., HOU, L., ZHANG, G. K., AND MOON, S. Development of bim, iot and ar/vr technologies for fire safety and upskilling. *Automation in Construction 125* (2021), 103631. doi:10.1016/j.autcon.2021.103631.

[13] CORREA, A., BARCELO, M., MORELL, A., AND VICARIO, J. L. A review of pedestrian indoor positioning systems for mass market applications. *Sensors 17*, 8 (2017). doi:10.3390/s17081927.

[14] CÔTÉ, M.-A., KÁDÁR, Á., YUAN, X., KYBARTAS, B., BARNES, T., FINE, E., MOORE, J., HAUSKNECHT, M., EL ASRI, L., ADADA, M., TAY, W., AND TRISCHLER, A. Textworld: A learning environment for text-based games. In *Computer Games* (Cham,

2019), T. Cazenave, A. Saffidine, and N. Sturtevant, Eds., Springer International Publishing, pp. 41–75. doi:10.1007/978-3-030-24337-1_3.

[15] DALTON, R., WILBERTZ, G., AND HOELSCHER, C. Perceptions of building-layout complexity. In *Spatial Cognition 2008 (SC '08), 15 September - 19 September 2008, Freiburg, Germany* (2008).

[16] DE COCK, L., OOMS, K., VAN DE WEGHE, N., AND DE MAEYER, P. Google indoor maps or Google indoor no maps? Usability study of an adapted mobile indoor wayfinding aid. In *HCI International 2020 - Posters* (Cham, 2020), C. Stephanidis and M. Antona, Eds., Springer International Publishing, pp. 413–420. doi:10.1007/978-3-030-50726-8_54.

[17] DE COCK, L., VAN DE WEGHE, N., OOMS, K., VANHAEREN, N., RIDOLFI, M., DE POORTER, E., AND DE MAEYER, P. Taking a closer look at indoor route guidance; usability study to compare an adapted and non-adapted mobile prototype. *Spatial Cognition & Computation 22*, 1-2 (2022), 51–73. doi:10.1080/13875868.2021.1885411.

[18] DUCKHAM, M., AND KULIK, L. "Simplest" paths: Automated route selection for navigation. In *Spatial Information Theory. Foundations of Geographic Information Science* (Berlin, Heidelberg, 2003), W. Kuhn, M. F. Worboys, and S. Timpf, Eds., Springer Berlin Heidelberg, pp. 169–185. doi:10.1007/978-3-540-39923-0_12.

[19] FLEISCHMANN, M. Momepy: Urban morphology measuring toolkit. *Journal of Open Source Software 4*, 43 (2019), 1807. doi:10.21105/joss.01807.

[20] FOGLIARONI, P., BUCHER, D., JANKOVIC, N., AND GIANNOPOULOS, I. Intersections of Our World. In *10th International Conference on Geographic Information Science (GIScience 2018)* (Dagstuhl, Germany, 2018), S. Winter, A. Griffin, and M. Sester, Eds., vol. 114 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, pp. 3:1–3:15. doi:10.4230/LIPIcs.GISCIENCE.2018.3.

[21] FORTUNE, S. A sweepline algorithm for Voronoi diagrams. In *Proceedings of the Second Annual Symposium on Computational Geometry* (New York, NY, USA, 1986), SCG '86, Association for Computing Machinery, p. 313–322. doi:10.1145/10515.10549.

[22] FRANK, A. U. Pragmatic information content—how to measure the information in a route. *Foundations of geographic information science* (2003), 47.

[23] FUNKE, J. F., SCHIKORR, A., KARAOSMANOGLU, S., HIRZLE, T., STEINICKE, F., AND RUKZIO, E. Tiles to move: Investigating tile-based locomotion for virtual reality. *Proc. ACM Hum.-Comput. Interact. 7*, CHI PLAY (oct 2023). doi:10.1145/3611060.

[24] GEDICKE, S., TOMKO, M., WINTER, S., AND HAUNERT, J.-H. Selecting landmarks for wayfinding assistance based on advance visibility. In *Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems* (New York, NY, USA, 2023), SIGSPATIAL '23, Association for Computing Machinery. doi:10.1145/3589132.3625605.

[25] GIBSON, J. J. *The Ecological Approach to Visual Perception*. Houghton Mifflin Company, Boston, 1979. doi:10.4324/9781315740218.

[26] GÖTZE, J., AND BOYE, J. *"Turn Left" Versus "Walk Towards the Café": When Relative Directions Work Better Than Landmarks*. Springer International Publishing, Cham, 2015, pp. 253–267. doi:10.1007/978-3-319-16787-9_15.

[27] GRASS DEVELOPMENT TEAM. *Geographic Resources Analysis Support System (GRASS GIS) Software, Version 8.2*. Open Source Geospatial Foundation, 2022.

[28] GUERREIRO, J. A., AHMETOVIC, D., SATO, D., KITANI, K., AND ASAKAWA, C. Airport accessibility and navigation assistance for people with visual impairments. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2019), CHI '19, Association for Computing Machinery, p. 1–14. doi:10.1145/3290605.3300246.

[29] HAGBERG, A., SWART, P. J., AND SCHULT, D. A. Exploring network structure, dynamics, and function using NetworkX. Tech. rep., Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2008. doi:10.25080/TCWV9851.

[30] HAMZEI, E., COCK, L. D., TOMKO, M., DE WEGHE, N. V., AND WINTER, S. Indoor view graph: A model to capture route and configurational information. *Environment and Planning B: Urban Analytics and City Science 0*, 0 (0), 23998083241241598. doi:10.1177/23998083241241598.

[31] HAQ, S., AND GIROTTO, S. Ability and intelligibility: Wayfinding and environmental cognition in the designed. In *4th international space syntax symposium, London* (2003), pp. 1–68.

[32] HAQUE, S., KULIK, L., AND KLIPPEL, A. Algorithms for reliable navigation and wayfinding. In *Spatial Cognition V Reasoning, Action, Interaction* (Berlin, Heidelberg, 2007), T. Barkowsky, M. Knauff, G. Ligozat, and D. R. Montello, Eds., Springer Berlin Heidelberg, pp. 308–326. doi:10.1007/978-3-540-75666-8_18.

[33] HÖLSCHER, C., AND DALTON, R. C. Comprehension of layout complexity: Effects of architectural expertise and mode of presentation. In *Design Computing and Cognition '08* (Dordrecht, 2008), J. S. Gero and A. K. Goel, Eds., Springer Netherlands, pp. 159–178. doi:10.1007/978-1-4020-8728-8_9.

[34] JANSEN, P. A. A systematic survey of text worlds as embodied natural language environments. *CoRR abs/2107.04132* (2021). doi:10.48550/arXiv.2107.04132.

[35] KLIPPEL, A. *Wayfinding choremes: Conceptualizing wayfinding and route direction elements*. University of Bremen, 2003. doi:10.1016/j.jvlc.2004.11.004.

[36] KLIPPEL, A., C., F., AND S., W. You-are-here maps in emergencies—the danger of getting lost. *Journal of Spatial Science 51*, 1 (2006), 117–131. doi:10.1080/14498596.2006.9635068.

[37] KLIPPEL, A., AND MONTELLO, D. R. Linguistic and nonlinguistic turn direction concepts. In *Spatial Information Theory* (Berlin, Heidelberg, 2007), S. Winter, M. Duckham, L. Kulik, and B. Kuipers, Eds., Springer Berlin Heidelberg, pp. 354–372. doi:10.1007/978-3-540-74788-8_22.

[38] KLIPPEL, A., TAPPE, H., KULIK, L., AND LEE, P. U. Wayfinding choremes—a language for modeling conceptual route knowledge. *Journal of Visual Languages & Computing 16*, 4 (2005), 311–329. doi:10.1016/j.jvlc.2004.11.004. Perception and ontologies in visual, virtual and geographic space.

[39] KLIPPEL, A., AND WINTER, S. Structural salience of landmarks for route directions. In *Spatial Information Theory* (Berlin, Heidelberg, 2005), A. G. Cohn and D. M. Mark, Eds., Springer Berlin Heidelberg, pp. 347–362. doi:10.1007/11556114_22.

[40] KRISP, J. M., AND KELER, A. Car navigation—computing routes that avoid complicated crossings. *International Journal of Geographical Information Science 29*, 11 (2015), 1988–2000. doi:10.1080/13658816.2015.1053485.

[41] KULIGA, S. F., NELLIGAN, B., DALTON, R. C., MARCHETTE, S., SHELTON, A. L., CARLSON, L., AND HÖLSCHER, C. Exploring individual differences and building complexity in wayfinding: The case of the seattle central library. *Environment and Behavior 51*, 5 (2019), 622–665. doi:10.1177/0013916519836149.

[42] KUNHOTH, J., KARKAR, A., AL-MAADEED, S., AND AL-ALI, A. Indoor positioning and wayfinding systems: a survey. *Human-centric Computing and Information Sciences 10*, 1 (May 2020), 18. doi:10.1186/s13673-020-00222-0.

[43] LI, R., AND KLIPPEL, A. Wayfinding behaviors in complex buildings: The impact of environmental legibility and familiarity. *Environment and Behavior 48*, 3 (2016), 482–510. doi:10.1177/0013916514550243.

[44] LIN LI, KAI MAO, G. L., AND WEN, Y. A landmark-based cognition strength grid model for indoor guidance. *Survey Review 50*, 361 (2018), 336–346. doi:10.1080/00396265.2016.1277004.

[45] LÜCKE, D., MOSSAKOWSKI, T., AND MORATZ, R. Streets to the OPRA—finding your destination with imprecise knowledge. In *IJCAI Workshop on Benchmarks and Applications of Spatial Reasoning. IJCAI Workshop on Benchmarks and Applications of Spatial Reasoning, located at IJCAI, July 17, Barcelona, Spain* (2011), J. Renz, A. G. Cohn, and S. Wölfl, Eds., pp. 25–32.

[46] MACMAHON, M., STANKIEWICZ, B., AND KUIPERS, B. Walk the talk: Connecting language, knowledge, and action in route instructions. *Def 2*, 6 (2006), 4.

[47] MARIE-PAULE, D., ARIANE, T., ELSA, M., AND DENIS, M. Testing the value of route directions through navigational performance. *Spatial Cognition & Computation 3*, 4 (2003), 269–289. doi:10.1207/s15427633scc0304_2.

[48] MARIE-PAULE, D., AND MICHEL, D. Spatial descriptions as navigational aids: A cognitive analysis of route directions. *Kognitionswissenschaft 7*, 1 (Feb 1998), 45–52. doi:10.1007/BF03354963.

[49] MORTARI, F., CLEMENTINI, E., ZLATANOVA, S., AND LIU, L. An indoor navigation model and its network extraction. *Applied Geomatics 11*, 4 (Dec 2019), 413–427. doi:10.1007/s12518-019-00273-8.

[50] MÜLLER, P., WONKA, P., HAEGLER, S., ULMER, A., AND VAN GOOL, L. Procedural modeling of buildings. In *ACM SIGGRAPH 2006 Papers* (New York, NY, USA, 2006), SIGGRAPH '06, Association for Computing Machinery, p. 614–623. doi:10.1145/1179352.1141931.

[51] NATAPOV, A., KULIGA, S., DALTON, R. C., AND HÖLSCHER, C. Building circulation typology and space syntax predictive measures. In *Proceedings of the 10th international space syntax symposium* (2015), vol. 12, Space Syntax Laboratory, The Bartlett School of Architecture, University ..., pp. 13–17. doi:10.1080/00038628.2019.1675041.

[52] NELSON, G. Afterword: Five years later. *IF Theory Reader* (2011), 189.

[53] NORMAN, D. A. Affordance, conventions, and design. *Interactions 6*, 3 (may 1999), 38–43. doi:10.1145/301153.301168.

[54] PAPPU, A., AND RUDNICKY, A. The structure and generality of spoken route instructions. In *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue* (2012), pp. 99–107.

[55] RICHTER, K.-F. Context-specific route directions. *Künstliche Intell. 22*, 3 (2008), 73–79.

[56] RICHTER, K.-F., WINTER, S., AND SANTOSA, S. Hierarchical representations of indoor spaces. *Environment and Planning B: Planning and Design 38*, 6 (2011), 1052–1070. doi:10.1068/b37057.

[57] RIESBECK, C. K. "You can't miss it": Judging the clarity of directions. *Cognitive Science 4*, 3 (1980), 285–303. doi:10.1207/s15516709cog0403_3.

[58] RUSSO, D., ZLATANOVA, S., AND CLEMENTINI, E. Route directions generation using visible landmarks. In *Proceedings of the Sixth ACM SIGSPATIAL International Workshop on Indoor Spatial Awareness* (New York, NY, USA, 2014), ISA '14, Association for Computing Machinery, p. 1–8. doi:10.1145/2676528.2676530.

[59] TAYLOR, E. G. R. Early charts and the origin of the compass rose. *Journal of Navigation 4*, 4 (1951), 351–356. doi:10.1017/S0373463300034214.

[60] TEIMOURI, F., AND RICHTER, K.-F. 'Straight? what straight?' investigating navigation instructions' applicability. *Journal of Location Based Services 17*, 1 (2023), 1–25. doi:10.1080/17489725.2021.2014582.

[61] TOMKO, M., AND RICHTER, K.-F. Defensive wayfinding: Incongruent information in route following. In *Spatial Information Theory* (Cham, 2015), S. I. Fabrikant, M. Raubal, M. Bertolotto, C. Davies, S. Freundschuh, and S. Bell, Eds., Springer International Publishing, pp. 426–446. doi:10.1007/978-3-319-23374-1_20.

[62] TOMKO, M., AND WINTER, S. Pragmatic construction of destination descriptions for urban environments. *Spatial Cognition & Computation 9*, 1 (2009), 1–29. doi:10.1080/13875860802427775.

[63] TOMKO, M., AND WINTER, S. Describing the functional spatial structure of urban environments. *Computers, Environment and Urban Systems 41* (2013), 177–187. doi:10.1016/j.compenvurbsys.2013.05.002.

[64] TSCHANDER, L. B., SCHMIDTKE, H. R., ESCHENBACH, C., HABEL, C., AND KULIK, L. A geometric agent following route instructions. In *Spatial Cognition III* (Berlin, Heidelberg, 2003), C. Freksa, W. Brauer, C. Habel, and K. F. Wender, Eds., Springer Berlin Heidelberg, pp. 89–111. doi:10.1007/3-540-45004-1_6.

[65] VANHAEREN, N., DE COCK, L., LAPON, L., VAN DE WEGHE, N., OOMS, K., AND DE MAEYER, P. On the right track: Comfort and confusion in indoor environments. *ISPRS International Journal of Geo-Information 9*, 2 (2020). doi:10.3390/ijgi9020132.

[66] WINTER, S., TOMKO, M., ELIAS, B., AND SESTER, M. Landmark hierarchies in context. *Environment and Planning B: Planning and Design 35*, 3 (2008), 381–398. doi:10.1068/b33106.

[67] YANG, L., AND WORBOYS, M. Similarities and differences between outdoor and indoor space from the perspective of navigation. *Poster presented at COSIT* (2011).

[68] ZAHABI, M., ZHENG, X., MAREDIA, A., AND SHAHINI, M. Design of navigation applications for people with disabilities: A review of literature and guideline formulation. *International Journal of Human–Computer Interaction 39*, 14 (2023), 2942–2964. doi:10.1080/10447318.2022.2088883.

[69] ZHAO, M., ANDERSON, P., JAIN, V., WANG, S., KU, A., BALDRIDGE, J., AND IE, E. On the evaluation of vision-and-language navigation instructions. *arXiv preprint arXiv:2101.10504* (2021).

[70] ZHICHENG ZHAN, HAOSHENG HUANG, S. W., AND DE WEGHE, N. V. Cognitive mapping of indoor environments: constructing an indoor navigation network from crowdsourced indoor route descriptions. *Cartography and Geographic Information Science 51*, 3 (2024), 385–403. doi:10.1080/15230406.2024.2305450.

8 Supplementary materials

valid route instruction	invalid route instruction
Welcome An Interactive Fiction Inform 7 build 6M62 (16/v6.33) An area (16) in r0 A door between a160 and a13r1 (first opening d1) You can enter the r1 by going north (at the left) You can continue in the r0 by going east (at the front) > Go straight An area (17) in r0 A door between a170 and a18r1 (first opening dO) You can enter the r1 by going north (at the left) You can continue in the re by going south (at the right) You can continue in the r0 by going west (at the back) X: 546472.9076679605 Y: 1000856.9294065498 > Turn right An area (14) in r0 You can continue in the r0 by going north (at the back) You can continue in the r0 by going east (at the left) You can continue in the r0 by going west (at the right) X: 546709.4310969368 Y: 1000167.4370029137 > Turn right Room 0 An area (7) in r0 You can continue in the r0 by going northwest (at the slight right) You can continue in the r0 by going southwest (at the slight left) You can continue in the r0 by going east (at the back) X: 546328.634469697 Y: 1000044.0018939395 > Turn slightly left Room 0 An area (6) in r0 You can continue in the r0 by going northeast (at the back) You can continue in the r0 by going southwest (at the front) X: 546312.2137161086 Y: 1000016.6339712921 > Go straight Room 0 An area (23) in r0 You can continue in the r0 by going northeast (at the back) X: 546281.0 Y: 999991.0 !!Valid Route Instruction, Intended Destination!!	Welcome An Interactive Fiction Inform 7 build 6M62 (I6/v6.33) Room 0 An area (16) in ro A door between a16r0 and a13r1 (first opening d1) You can enter the r1 by going north (at the left) You can continue in the r0 by going east (at the front) > Go straight Room 0 An area (17) in r0 A door between a17r0 and a18r0 (first opening do) You can enter the r1 by going north (at the left) You can continue in the r0 by going south (at the right) You can continue in the r0 by going west (at the back) X: 546472.9076679605 Y: 1000856.9294065498 > Turn right An area (14) in re You can continue in the r0 by going north (at the back) You can continue in the r0 by going east (at the left) You can continue in the r0 by going west (at the right) X: 546709.4310969368 Y: 1000167.4370029137 > Turn right An area (7) in r0 You can continue in the r0 by going northwest (at the slight right) You can continue in the r0 by going southwest (at the slight left) You can continue in the r0 by going east (at the back) X: 546328.634469697 Y: 1000044.0018939395 > Turn left You can't go that way. X: 546328.634469697 Y: 1000044.0018939395 > Go straight You can't go that way. X: 546328.634469697 Y: 1000044.0018939395 !! Invalid Route Instruction, Wrong Destination!!

Table S7: Valid and invalid route instructions in action in a simulated indoor environment in a text-based game