GAIA: A Givenness Hierarchy Theoretic Model of Situated Referring Expression Generation

Mark Higger (mhigger@mines.edu)

Tom Williams (twilliams@mines.edu)

MIRRORLab, Colorado School of Mines
Golden, CO, USA

Abstract

A key task in natural language generation (NLG) is Referring Expression Generation (REG), in which a set of properties are selected to describe a target referent. Computational cognitive models of REG typically focus on REG-in-context, where the referring expressions are designed to take into account the conversational context into which they are to be generated. However, in practice, these methods only focus on *linguistic* context of the *text* into which they are to be inserted. We argue that to develop robust models of naturalistic human referring, REG will need to move beyond linguistic context, and account for cognitive and environmental context as well. That is, we propose that a cognitivist, interactionist, and situated approach to modeling REG is needed. In this paper, we present GAIA, a Givenness Hierarchy theoretic model of REG, and demonstrate the immediate qualitative benefits of this model over the traditional REG model which it extends.

Keywords: Natural Language Generation; Referring Expression Generation; Cognitive Status; Givenness Hierarchy

Introduction

Computationally modeling how humans generate natural language utterances is a key task both for those pursuing *natural language generation as science* (Deemter, 2023) and for those aiming to engineer effective natural language generation systems (Reiter & Dale, 1997). For researchers of both stripes, modeling human processes for Referring Expression Generation (REG) (in which speakers select the properties they will use to refer to a target referent) has stood as a key subtask (Van Deemter, 2016).

Recently, work in the field of Referring Expression Generation has begun to move from one-shot REG (Krahmer & Van Deemter, 2012), in which a set of properties are selected to disambiguate a target referent with respect to a set of distractors, to REG-in-Context (Belz & Varges, 2007; Chen et al., 2023), where features of the dialogue state are used to inform the selection of properties. For example, a computational cognitive model of REG-in-Context might demonstrate its ability to model humanlike REG by taking a series of *unfilled* utterances like those seen in the following *Task Example*, and translate them into *filled* utterances in the following *Solution Example*.

Context — A speaker at the front desk of a hospital witnesses a person put non-recyclable trash into a recycling bin [b1] nearest to the desk. The speaker wishes to notify the person to put the trash into a trash bin instead.

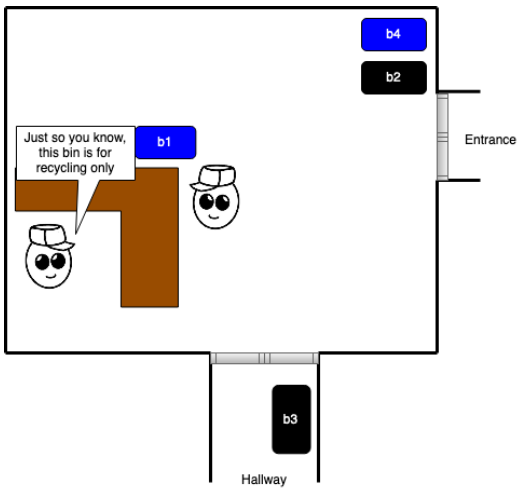

Figure 1: Visualization of motivating context.

Task Example —

1. "Hi, just so you know, [b1] is for recycling only"
2. "Please use [b2], which is for all trash"
3. "If [b2] is full, you can also use [b3]"

Solution Example —

1. "Hi, just so you know, [this bin] is for recycling only"
2. "Please use [the black bin by the entrance], which is for all trash"
3. "If [that bin] is full, you can also use [the bin in the hallway]"

Computational cognitive models of *REG-in-Context* typically break this task into three key steps (cf. Van Deemter, 2016; Levelt et al., 1999): (1) Referring Form Selection (RFS), in which a referring form such as '*it*', '*this*', '*that*-$\langle NP' \rangle$', or $\langle$*proper-noun*$\rangle$ is chosen (Kibrik, 2011; Han & Williams, 2023; Del Castillo et al., 2023); (2) Content Selection, in which, *if* a noun-phrase bearing Referring Form was selected (e.g., '*this*-$\langle NP' \rangle$', '*that*-$\langle NP' \rangle$', '*the*-$\langle NP' \rangle$'), the properties to be used for that NP are selected (e.g., $black(X)$, $bin(X)$) (Van Deemter, 2016; Dale, 1989; Reiter, 1990); (3) Content Realization, in which the properties are translated into specific words that communicate those properties (Mal-

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

ouf, 2000; Mitchell et al., 2011).

While Content Realization has been historically viewed as a separate task from REG, we list it here due to the recent rise of Neural REG Systems (Chen et al., 2023; Ferreira et al., 2018; Cao & Cheung, 2019; Cunha et al., 2020; Same et al., 2022), which generate referring expressions in an end-to-end manner without breaking the problem into these constituent parts. These methods have attracted significant recent attention due in part to the general rise of the use of transformer networks (Jaderberg et al., 2015) based Large Language Models (LLMs), such as chatGPT (OpenAI, 2024), which are capable of simultaneous linguistic realization and content determination that generally aligns with the patterns of natural human speech – at least as it manifests in the web corpora on which it is trained (Dathathri et al., 2019). While these neural methods have achieved some success from an engineering standpoint, they fail to serve as useful computational *cognitive* models for several key reasons. As Chen et al. (2023) analyze, these models are uninterpretable, tend to be trained on web-based text corpora that do not represent natural human-like speech, and they focus narrowly on realizing the surface features of Western European languages – and, we would add, of the Standardized White variants of those languages. For these and other reasons, these models fail to serve as generalizable cognitive models that shed insights into the mechanisms and representations used in human language generation at a cognitivist level of analysis.

Moreover, because these types of models for REG-in-Context focus on generating *texts*, the notion of "context" in these works is limited to the *linguistic context* of the ongoing text generation task. In contrast, we argue that if we truly wish to understand the mechanisms of human language production, we need to take a stance that is not only *cognitivist* (that is, focusing on and elucidating of the mental representations and cognitive processes of language generation), but also *interactionist* (that is, focusing on the ways that language generation is performed *in relation to* other specific social agents with whom one is interacting, and who may have a different understanding and awareness of the world around them), and *situated* (that is, focusing on the ways that those interpersonal interactions are embedded into a specific spatial environment with key referentially-relevant dimensions such as proximity and visibility (Han & Williams, 2023)).

Recently, cognitive scientists have begun to take a cognitivist, interactionist, and situated approach to RFS, in two key ways. First, researchers have begun to integrate models of *cognitive status* into RFS models, so that referring forms account for whether the social agent with whom the speaker is interacting is already focusing on, or otherwise attending to, the referent in question (which allows the use of referring forms like '*it*' or '*that*') (Han & Williams, 2023). Second, researchers have begun to incorporate knowledge of situated features such as referent distance into RFS model (allowing discriminating use of forms like '*this*' vs. '*that*').

We argue in this work that a similar movement needs to be taken with respect to the Content Selection stage of REG, which has long been regarded as the key REG task. As we will show in this work, integrating models of *cognitive status* into cognitivist REG Content Selection algorithms fundamentally and qualitatively improves the outputs of those models in situated interaction contexts, especially when those cognitive status models themselves account for features of the environmental context in which the interaction is embedded. Specifically, we present a novel REG Algorithm, the Givenness-Advised Incremental Algorithm (GAIA), that uses these context-sensitive cognitive status models to generate Referring Expressions in a way that is dramatically more efficient and natural. GAIA achieves these gains by eliminating ostensible "distractors" that fall outside the bounds of what listeners would likely find relevant, given the cognitive status cues made by the speaker's choice of referring form.

In this paper we will first further motivate this work through exploration of related literature. We will then present GAIA, and provide an algorithm walkthrough. Then, we will use the scenario described above as a case study, showing how GAIA's operation within this case study clearly demonstrates qualitative and advantageous differences between GAIA and classic REG models. Finally, we will discuss the limitations that bound these advantages, and suggest directions for future work.

Related Work

Computational Models of Referring

Early computational cognitive models of referring largely took inspiration from the Gricean Maxims (Grice, 1975). These maxims, while notoriously underformalized (Van Deemter, 2016), make claims to the nature of typical cooperative conversation, including how much is said (quantity), how truthful speakers are (quality), how relevant speakers are (relation), and how clear speakers strive to be (manner) (Grice, 1975). The early Full Brevity (FB) and Greedy Algorithm (GR) REG methods specifically aimed to generate utterances that maximally adhered to the maxims of quantity and manner, by crafting referring expressions that were as brief as possible while being fully unambiguous (Dale, 1989; Dale & Reiter, 1995). One challenge faced by these methods was that humans do not always follow these conversational maxims. For example, humans do not always generate minimal descriptions, due in part for their latent preferences for easy-to-process properties, and due in part to the incremental nature of reference production (Pechmann, 1989). To address these caveats, Reiter & Dale (1997) introduced the Incremental Algorithm (IA), which operates by incrementally considering properties according to a preference ordering, and adopting those properties to rule out distractors. Due to its simplicity and effectiveness, this algorithm is still considered the standard for REG a quarter-century since its introduction (Van Deemter, 2016).

Despite its popularity, the simplicity of the IA belies a number of flaws, especially within the interactive, situated

contexts emphasized in this work. The IA is unable to account for uncertainty, is unable to generate relations between entities, as well as assumptions about the ways information about the entity is stored. Yet most significantly, the IA assumes that all entities being considered are equally relevant, and thus depends on a well-defined scope of entities to consider. While these assumptions are not unreasonable for traditional REG research conducted within purely textual domains, they are unrealistic both for modeling human cognition, as well as in non-textual engineering domains like robotics. As a result, researchers working at the intersection of cognitive science and robotics have designed algorithms like *DIST-PIA* (Williams & Scheutz, 2017; Williams, Thielstrom, et al., 2018), which simultaneously serve as more apt computational cognitive models, and as more practical engineering solutions for robotic domains.

Even these recent algorithms, however, do not define the scope of entities to consider and do not account for *cognitive* context, nor for the ways *environmental* context shapes that cognitive context. As an example, let us briefly reconsider the example scenario introduced above. In this context, the speaker and hearer know of at least four bins (and in fact, they may well know of dozens more). Yet, as shown in the Solution Example, the speaker can regularly use expressions (e.g., "this bin") whose properties alone would fail to fully disambiguate the target referent. Nevertheless, the speaker can confidently use those underdetermined demonstratives (Clark et al., 1983), because they know that the entities they are referring to will be sufficiently disambiguated *with respect to the entities their interlocutor will believe to be sufficiently relevant to the conversation, on the basis of the common ground they share with their interlocutor about their shared environmental context* (cf. Clark et al., 1983). One key linguistic framework for reasoning about how these subsets of relevant entities are delineated in conversation is the Givenness Hierarchy.

Givenness Hierarchy Theoretic Computational Cognitive Modeling

Presented by Gundel et al. (1993), the Givenness Hierarchy argues that pieces of information hold different tiers of cognitive status in the human mind; that speakers consciously or subconsciously reason about the status information holds in the minds of their interlocutors; that speakers' choice of referring form signals the cognitive status of the information to which they intend to refer; and that listeners use these cues to circumscribe the set of possible referrents of speakers' referring expressions. Specifically, a piece of information can be said to have one of the following six cognitive statuses:

1. In Focus: The entity is at the center of attention
2. Activated: The entity is represented in working memory, but is not necessarily the center of attention.
3. Familiar: Entity is represented in memory, while not necessarily being represented in working memory.
4. Uniquely Identifiable: Entity can be accessed uniquely, without necessarily being represented in memory
5. Referential: Entity can be accessed, but not necessarily accessed uniquely
6. Type Identifiable: The type of entity can be accessed, but not necessarily an instance of the entity

Critically, these cognitive statuses are *hierarchical* in that all entities that are of a particular cognitive status also can be said to have all lower cognitive statuses. For example, an 'Activated' object is also 'Familiar', 'Uniquely Identifiable', 'Referential', and 'Type Identifiable'.

Each of these tiers of cognitive status is then associated with a different set of referring forms: An entity that is at least uniquely identifiable can be described using the referring form '$the\langle N' \rangle$'. An entity that is at least familiar can also be described using the referring form '$that\langle N' \rangle$'. An entity that is at least activated can also be described using the referring form '*this*', '*that*', or '$this\langle N' \rangle$'. And an entity that is in focus can be described using '*it*'.

Inversely, then, each of these referring forms can be used to indicate a lower bound on the cognitive status the speaker assumes their referent has in the mind of the speaker, and *thus specifies a context set of relevant entities held in ground with their interlocutor*. If a speaker uses "this bin" to refer to an object, the listener can assume that the speaker is referring to an entity that is in their working memory, and for which it is reasonably likely that the speaker and hearer hold *as a matter of common ground* (Clark & Carlson, 1981) that the listener is thinking about the bin at that level. If the listener makes this inference, it need not consider all possible bins when resolving the speaker's reference; they need only consider those entities that hold that status. As a second example, if a speaker uses 'it', the listener can assume that the object being referred to is in their focus of attention, or more specifically, is the object of their joint attention. Recently, both roboticists and cognitive scientists have used this theoretical intuition to develop better models of human and robot language understanding and generation.

In the language understanding literature, several models of Givenness Hierarchy theoretic *reference resolution* have been developed over the past twenty years (Chai et al., 2006; Williams et al., 2016; Williams & Scheutz, 2019; Williams, Krause, et al., 2018). Of particular note, Williams' GH-POWER (Williams & Scheutz, 2019) and GROWLER (Williams, Krause, et al., 2018) algorithms maintain second-order theory of mind models, using linguistically informed rules to maintain buffers containing what entities the listener estimates that the speaker might assume the listener to assume to have different cognitive statuses. During reference resolution the listener then must only disambiguate between the entities contained in the data structures associated with the cognitive status cued by the speaker's choice of referring form (and those of higher statuses).

More recently, the Givenness Hierarchy has also been

leveraged in Natural Language Generation both for "document planning" (i.e., pre-planning the sequences of utterances one will use when communicating a multi-step task) (Spevak et al., 2022)) and for referring form selection (Han & Williams, 2023; Del Castillo et al., 2023). In the latter, the speaker recursively estimates the cognitive status of task-relevant entities in the mind of their interlocutor, and then uses this information *in combination* with environmentally-relevant features such as the distance to the target referent, to decide which referring form to use.

We believe there is a key opportunity to similarly use this framework when selecting the content of referring expressions. Once the speaker has selected a referring form on the basis of its presumed cognitive status, (assuming the selected referring form includes a noun phrase) they may more intelligently select properties to include in their referring expression if they assume that their listener will correctly infer an appropriately bounded context set as indicated by that referring form. Doing so will allow the speaker to select a smaller number of properties, as they must only select properties that eliminate distractors in that reduced context set.

In the following section, we present the Givenness-Advised Incremental Algorithm (GAIA), a modified version of the Incremental Algorithm that adopts exactly this intuition. We choose to directly extend the IA rather than more complex algorithms like DIST-PIA in order to most cleanly demonstrate the advantages of this insight. As we will later discuss, the insights borne by GAIA and by DIST-PIA could easily be combined in future work to yield a more comprehensive computational cognitive model.

Algorithm and Walkthrough

In this section, we present the Givenness-Advised Incremental Algorithm (GAIA), a computational cognitive model of reference that extends the Incremental Algorithm (IA) (Dale & Reiter, 1995) by leveraging the Givenness Hierarchy (Gundel et al., 1993) to better address larger scale environments in real-world settings. Specifically, GAIA operates by proactively eliminating distractors that have a cognitive status lower than that of the target referent before beginning the REG process as conceptualized by the IA. This modification represents a simple yet tractable way to significantly reduce the number of irrelevant distractors that would otherwise be needed to be assessed, and to thus reduce the number of properties needed to formulate a unique description of the target referent. Because the *way* GAIA performs this initial distractor elimination is grounded in cognitive status, which is itself tightly tied to dialogue context, this approach naturally addresses the challenge of repeated reference. Similarly, because cognitive status is also tightly tied to environmental context, this approach naturally addresses challenges that arise from the environmental structure of large-scale interaction contexts.

Let us now walk carefully through GAIA. Similarly to the IA, GAIA starts by initializing a list of distractors (X) to all known entities (M), excepting the target entity (m_t) (Line 1). Unlike the IA, however, GAIA takes an additional step before proceeding. Specifically, GAIA considers the cognitive status

Notation

- D Incrementally built up list of descriptors
- P Queue of all properties in preference order consisting of $\{p_0, ..., p_i\}$
- M Model of all entities in the environment consisting of $\{m_0, ..., m_n\}$, where each entity contains values for each property (v_{pi})
- m_t Target entity for referring expression
- c_m Cognitive status for entity m
- $v_{m,p}$ Property p value for entity m
- X Incrementally pruned set of distractors

Algorithm 1 GAIA: Givenness-Advised Incremental Algorithm

1: $X = M$ / m_t // Set of distractors equal to all entities except target referent
2: $c_t = m_t[c]$ // Get cognitive status of target referent
3: // Remove all distractors who's cognitive status is lower than the target referent
4: **for** x in X **do**
5: $c_x = x[c]$
6: **if** $c_x < c_t$ **then**
7: $X = pop(x)$
8: **end if**
9: **end for**
10: // Use the incremental algorithm to find the description using the remaining distractors
11: D = new Queue() // Initialize the Description
12: **while** $X \neq \emptyset$ and $P \neq \emptyset$ **do**
13: // For each property in preference order, find the new set of potential distractors
14: $p = pop(P)$
15: $v_{m_t,p} = m_t[p]$
16: $X' = \emptyset$
17: **for** $x = pop(X)$ **do**
18: // Add to the new distractor list any entity who has the same property value as the target referent
19: $v_{x,p} = x[p]$
20: **if** $v_{m_t,p} == v_{x,p}$ **then**
21: $X' = push(x)$
22: **end if**
23: // Only add the property value to the description if the new distractor list is smaller than the old one
24: **if** $X' \neq X$ **then**
25: $D = push(v_{m_t,p})$
26: $X = X'$
27: **end if**
28: **end for**
29: **end while**
30: **return** D

of the target (c_t) (Line 2), and then proactively eliminates all distractors whose cognitive status is lower (i.e., less restrictive) than c_t (Lines 2– 9). That is, if the target referent has a cognitive status of 'Activated', then an entity with a cognitive status of 'Familiar', 'Uniquely Identifiable', 'Referential', or 'Type Identifiable' would be immediately eliminated from the distractor list, but an entity with a cognitive status of 'Activated' or 'In Focus' would not.

The rest of GAIA then directly follows the procedure of IA. Once the initial distractors (X) have been set, an empty description (D) of the target referent is created (Line 11). All properties (P) are then iterated through to eliminate distractors in preference order until either there are no remaining distractors or all properties have been iterated. Then for each property, the corresponding property for the target referent ($v_{m_t,p}$) is found (Line 15). All entities in the previous list of distractors are iterated through, and the new list of distractors is populated with all of the previous distractors whose property values are the same as the target referent. Finally, if any distractors were ruled out during this process, the considered property is added to the description of the target referent (Lines 24– 27).

Case Study Validation

Case Study Definition

To demonstrate the benefits of GAIA, we provided our example scenario (depicted in Fig. 1) to a computational implementation of both GAIA and the IA, and contrasted the model outputs as a case study. Specifically, we first created a knowledge base containing knowledge representations for all bins present, with information about their color, location, and initial cognitive status. Second, we identified the key reference points in each utterance of the dialogue, and associated each with a target referent within the interaction context.

1. "Hi, just so you know, [b1] is for recycling only"
2. "Please use [b2], which is for all trash"
3. "If [b2] is full, you can also use [b3]"

We then proceeded through each reference point in each utterance in the example dialogue. At each reference point, we provided the current knowledge base to the IA and GAIA, retrieved the referring expressions generated by each algorithm, and then updated the cognitive status of each entity within the knowledge base. The method for updating cognitive status is described in the next section.

Cognitive Status Dynamics Policy

In order to use GAIA for REG, it is required to know the cognitive status of every entity. To do this there are many methods available to compute cognitive status. For example Pal et al. (2020) proposes a probabilistic Bayesian model to recursively estimate an entity's cognitive status based on verbal features. Alternatively, Spevak et al. (2022) uses a rules-based approach to estimate entities' cognitive statuses, however, this is similarly limited to only linguistic features. While GAIA can be effectively used in purely linguistic contexts, the presented case study is specifically chosen to highlight that it can also be used in multi-modal settings. While some recent research has proposed preliminary probabilistic computational multi-modal cognitive status estimation (Daigler et al., 2024), no such model has yet been implemented.

With that in mind, to determine the cognitive status, we use a rules-based estimator to determine what the cognitive status should be for each entity at each instance based on the cognitive status coding criteria outlined in Gundel et al. (2006). While these rules can be implemented computationally within the context of cognitive architectures such as DIARC (Scheutz et al., 2019), ACT-R (Anderson et al., 1997), or SOAR (Laird, 2019), we have chosen to manually code cognitive status to alleviate any errors that may arise from the estimator itself. Specifically, we use the following rules to determine cognitive status:

1. An entity is considered at least 'Uniquely Identifiable' if the speaker has enough properties in the entity representation to uniquely identify it.
2. An entity is considered at least 'Familiar' if both interactants have a representation of the entity in memory. For this scenario, we can assume this is true if the entity is in the room or has been mentioned in the conversation.
3. An entity is considered at least 'Activated' if it has previously been mentioned or gestured to recently.
4. An entity is considered 'In Focus' if it was mentioned in the topic role of the last utterance.

Using this scheme, the task objects have the following cognitive statuses at the start of the experiment: All bins are at least 'Uniquely Identifiable' since the speaker has a unique representation of them. 'b1', 'b2', and 'b4', are at least 'Familiar' since they are all in the same room as the interaction. Conversely, 'b3' is not considered 'Familiar' because the listener has not seen the object before, and thus does not have a representation in memory. Then, only 'b1' is considered at least 'Activated' since the person is throwing away trash directly into the bin just prior to the interaction.

Case Study Walkthrough

We are now ready to walk through our case study step by step, the results of which are shown in Table 1.

Reference Point 1 — The first utterance contained one reference point, at which a referring expression for $b1$ was requested to the IA and GAIA. In response, the IA selected properties $\{bin(b1), blue(b1), by-desk(b1)\}$ (i.e. using all possible properties), whereas GAIA selected properties $\{bin(b1)\}$ (i.e., acknowledging that no other properties were required to discriminate $b1$, as it was already at least activated and no other entities were at least activated). After this first reference point, $b1$ becomes 'In Focus' as it is the topic of utterance 1.

Entity Representation	IA	GAIA
"Hi, just so you know, [b1] is for recycling only"	"Hi, just so you know, {***bin(b1), blue(b1), by-desk(b1)***} is for recycling only"	"Hi, just so you know, {***bin(b1)***} is for recycling only"
"Please use [b2], which is for all trash"	"Please use {***bin(b2), black(b2), by-entrance(b2)***}, which is for all trash"	"Please use {***bin(b2), black(b2), by-entrance(b2)***}, which is for all trash"
"If [b2] is full, you can also use [b3]"	"If {***bin(b2), black(b2), by-entrance(b2)***} is full, you can also use {***bin(b3), in-hallway(b3)***}"	"If {***bin(b2)***} is full, you can also use {***bin(b3), in-hallway(b3)***}"

Table 1: Evaluation of different REG algorithms for scenario posed in Figure 1.

Reference Point 2 — The second utterance contains one reference point, at which a referring expression for $b2$ was requested to the IA and GAIA. In response both the IA and GAIA selected properties $\{bin(b2), blue(b2), by-entrance(b2)\}$ (i.e. using all possible properties of the bin) as the properties needed to discriminate $b2$ from all task objects ($b1, b3, b4$) were identical to the properties needed to discriminate from all familiar task objects ($b1, b4$). After the second reference point, $b1$ loses the 'In Focus' status, as it is not mentioned in the topic role of this sentence, however, it does stay 'Activated' as it was mentioned in the previous utterance. Conversely, $b2$ becomes both 'Activated' and 'In Focus' as it is mentioned in the topic role of this utterance.

Reference Point 3 — The third utterance contains 2 reference points, the first of which a referring expression for $b2$ was requested to the IA and GAIA. In response the IA selected properties $\{bin(b2), black(b2), by-entrance(b2)\}$ (i.e. using all possible properties of the bin) while GAIA selected properties $\{bin(b2)\}$ (i.e., acknowledging that no other properties were required to discriminate $b2$, as it was the only entity that was 'In-Focus'). After this reference point, there is no change in cognitive statuses for any of the entities.

Reference Point 4 — For the second reference point in utterance three, $b3$ was requested to the IA and GAIA. In response, both the IA and GAIA select the properties $\{bin(b3), in-hallway(b3)\}$ (i.e. using only the necessary properties of the bin to uniquely identify it from all other task objects).

Discussion

At reference point 1 GAIA only chooses the property of $\{bin(b1)\}$ despite the IA needing more properties to identify $b1$. This expression highlights one of the primary advantages of using a cognitive context over purely linguistic context. This is because it allows us to take into account the cognitive context that $b1$ is a direct part of the conversation, despite not being mentioned. This is important because in other REG-in-context algorithms have no way to account for this type of implicit context that is defined by action and environment rather than by dialog directly.

At reference point 2, GAIA only considers task objects that are at least familiar ($b1$, $b4$), while the IA considers all task objects ($b1$, $b3$, $b4$). However, since GAIA already needs to use both color and location of $b2$ to distinguish it from $b1$ and $b4$, the chosen properties also distinguish it from $b3$ causing an identical referring expression in both the IA and GAIA. This reference point highlights that while the IA and GAIA can produce identical referring expressions, by removing distractors before iterating through entity properties, accounting for cognitive context increases the computational efficiency for large-scale environments.

At reference point 3, after entity $b2$ is repeated, GAIA produces a reference which only uses the property $\{bin(b2)\}$. In contrast, the IA generates the exact same expression for $b2$ as it did in reference point 2, despite the repeated reference. This is one of the most common criticisms of IA, which is its inability to innately handle REG-in-context. With this reference point, we highlight that by simply accounting for cognitive context at an entity level we can easily extend the IA to achieve REG-in-context.

At reference point 4, both GAIA and the IA select the properties $\{bin(b3), in-hallway(b3)\}$. Notably, the property $color(x)$ is not present in this output. This is because when the appropriate context set is delineated by the speaker's choice of referring form, $b3$ can be distinguished by its location only. Note also here that this approach is able to generate a natural description for the bin in the hallway which, although the speaker cannot assume the listener knows about, can be assumed to be uniquely identifiable through definite reference.

Conclusion

In this paper, we highlight the need to account for cognitive and environmental context in REG. To address this identified need, we present GAIA, an extension of the IA that leverages the Givenness Hierarchy to enable a cognitivist, interactionist, situated approach. Specifically, we demonstrate how the cognitive status of entities can be used to define the search space used for REG. While this work does not aim to quantify the magnitude of the benefit provided by GAIA, our case study qualitatively demonstrates the advantages of our approach, and shows why cognitive context is important for REG. Overall, our work provides a simple, theoretically well-grounded, and clearly motivated approach toward situated language generation that accounts for *environmental* and *cognitive* context.

Acknowledgements

This work has been supported in part by the Office of Naval Research grant N00014-21-1-2418.

References

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). Act-r: A theory of higher level cognition and its relation to visual attention. *Human–Computer Interaction*, *12*(4), 439–462.

Belz, A., & Varges, S. (2007). Generation of repeated references to discourse entities. In *Proceedings of the eleventh european workshop on natural language generation (enlg 07)* (pp. 9–16).

Cao, M., & Cheung, J. C. K. (2019). Referring expression generation using entity profiles. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3163–3172).

Chai, J. Y., Prasov, Z., & Qu, S. (2006). Cognitive principles in robust multimodal interpretation. *Journal of Artificial Intelligence Research*, *27*, 55–83.

Chen, G., Same, F., & van Deemter, K. (2023). Neural referential form selection: Generalisability and interpretability. *Computer Speech & Language*, *79*, 101466.

Clark, H. H., & Carlson, T. B. (1981). Context for comprehension. *Attention and performance IX*, *313*, 30.

Clark, H. H., Schreuder, R., & Buttrick, S. (1983). Common ground at the understanding of demonstrative reference. *Journal of verbal learning and verbal behavior*, *22*(2), 245–258.

Cunha, R., Ferreira, T. C., Pagano, A., & Alves, F. (2020). Referring to what you know and do not know: Making referring expression generation models generalize to unseen entities. In *Proceedings of the 28th international conference on computational linguistics* (pp. 2261–2272).

Daigler, L., Higger, M., Mott, T., & Williams, T. (2024). Challenges in annotating gesture-based cognitive status in human-robot collaboration datasets. In *19th annual acm/ieee international conference on human robot interaction (hri).*

Dale, R. (1989). Cooking up referring expressions. In *27th annual meeting of the association for computational linguistics* (pp. 68–75).

Dale, R., & Reiter, E. (1995). Computational interpretations of the gricean maxims in the generation of referring expressions. *Cognitive science*, *18*(2), 233–263.

Dathathri, S., Madotto, A., Lan, J., Hung, J., Frank, E., Molino, P., ... Liu, R. (2019). Plug and play language models: A simple approach to controlled text generation. *arXiv preprint arXiv:1912.02164*.

Deemter, K. v. (2023). Dimensions of explanatory value in nlp models. *Computational Linguistics*, *49*(3), 749–761.

Del Castillo, G., Clark, G., Han, Z., & Williams, T. (2023). Exploring the naturalness of cognitive status-informed referring form selection models. In *Proceedings of the 16th international natural language generation conference* (pp. 269–278).

Ferreira, T. C., Moussallem, D., Kádár, A., Wubben, S., & Krahmer, E. (2018). Neuralreg: An end-to-end approach to referring expression generation. *arXiv preprint arXiv:1805.08093*.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Gundel, J. K., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 274–307.

Gundel, J. K., Hedberg, N., Zacharski, R., Mulkern, A., Custis, T., Swierzbin, B., ... others (2006). Coding protocol for statuses on the givenness hierarchy. *Unpublished manuscript (1993/2006). http://www. sfu. ca/hedberg/Coding_ for_Cognitive_Status. pdf*.

Han, Z., & Williams, T. (2023). Evaluating cognitive status-informed referring form selection for human-robot interactions. In *2023 annual meeting of the cognitive science society (cogsci).*

Jaderberg, M., Simonyan, K., Zisserman, A., et al. (2015). Spatial transformer networks. *Advances in neural information processing systems*, *28*.

Kibrik, A. A. (2011). *Reference in discourse*. Oxford Studies in Typology and.

Krahmer, E., & Van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Computational Linguistics*, *38*(1), 173–218.

Laird, J. E. (2019). *The soar cognitive architecture*. MIT press.

Levelt, W. J., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and brain sciences*, *22*(1), 1–38.

Malouf, R. (2000). The order of prenominal adjectives in natural language generation. In *Proceedings of the 38th annual meeting of the association for computational linguistics* (pp. 85–92).

Mitchell, M., Dunlop, A., & Roark, B. (2011). Semi-supervised modeling for prenominal modifier ordering. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies* (pp. 236–241).

OpenAI. (2024). *Chatgpt.* Retrieved from `https://chat.openai.com/`

Pal, P., Zhu, L., Golden-Lasher, A., Swaminathan, A., & Williams, T. (2020). Givenness hierarchy theoretic cognitive status filtering. *arXiv preprint arXiv:2005.11267*.

Pechmann, T. (1989). Incremental speech production and referential overspecification. *Linguistics*.

Reiter, E. (1990). The computational complexity of avoiding conversational implicatures. In *28th annual meeting of the association for computational linguistics* (pp. 97–104).

Reiter, E., & Dale, R. (1997). Building applied natural language generation systems. *Natural Language Engineering*, *3*(1), 57–87.

Same, F., Chen, G., & van Deemter, K. (2022). Non-neural models matter: a re-evaluation of neural referring expression generation systems. In *Proceedings of the 60th annual meeting of the association for computational linguistics* (pp. 5554–5567).

Scheutz, M., Williams, T., Krause, E., Oosterveld, B., Sarathy, V., & Frasca, T. (2019). An overview of the distributed integrated cognition affect and reflection diarc architecture. *Cognitive architectures*, 165–193.

Spevak, K., Han, Z., Williams, T., & Dantam, N. T. (2022). Givenness hierarchy informed optimal document planning for situated human-robot interaction. In *2022 ieee/rsj international conference on intelligent robots and systems (iros)* (pp. 6109–6115).

Van Deemter, K. (2016). *Computational models of referring: a study in cognitive science*. MIT Press.

Williams, T., Acharya, S., Schreitter, S., & Scheutz, M. (2016). Situated open world reference resolution for human-robot dialogue. In *2016 11th acm/ieee international conference on human-robot interaction (hri)* (pp. 311–318).

Williams, T., Krause, E., Oosterveld, B., & Scheutz, M. (2018). Towards givenness and relevance-theoretic open world reference resolution. In *Rss workshop on models and representations for natural human-robot communication.*

Williams, T., & Scheutz, M. (2017). Referring expression generation under uncertainty: Algorithm and evaluation framework. In *Proceedings of the 10th international conference on natural language generation* (pp. 75–84).

Williams, T., & Scheutz, M. (2019). Reference in robotics: A givenness hierarchy theoretic approach. *The Oxford handbook of reference*.

Williams, T., Thielstrom, R., Krause, E., Oosterveld, B., & Scheutz, M. (2018). Augmenting robot knowledge consultants with distributed short term memory. In *Social robotics: 10th international conference, icsr 2018, qingdao, china, november 28-30, 2018, proceedings 10* (pp. 170–180).