Interpreting Vague and Ambiguous Referring Expressions by Dynamically Binding to Properties of the Context Set

Dustin A. Smith and Henry Lieberman⋆

MIT Media Lab; 20 Ames St. Cambridge, MA 02139 USA
{dustin,lieber}@media.mit.edu

Abstract. Referring expressions with vague and ambiguous modifiers, such as "a quick visit" and "the big meeting," are difficult for computers to interpret because their words' meanings are in part defined by context, which changes throughout the course of an interpretation. In this paper, we present an approach to interpreting context-dependent referring expressions that uses *dynamic binding*. During the incremental interpretation of a referring expression, a word's meaning can be defined in part by properties from the current candidate referents—its denotation up to the previous word for the tentative interpretation.

1 Referring expressions in context

For a hearer to understand the *intended meaning* of a speaker's utterance, he must make inferences based not only on evidence in the utterance's linguistically-encoded *surface meaning* but also on outside information, collectively referred to as **discourse context** (for an introduction, read [1, Ch. 1]). Nowhere is this more evident than in linguistic *reference*—when a speaker attempts to use her utterance to convey the identity of some entities (or set of entities) to her audience. The speaker does so by producing a **referring expression**, namely: "a description of an entity [or entities] that enables the hearer to identify that entity in a given context" [2]. Consider the referring expression "it": it seems evident that for the hearer to resolve what meaning the speaker intended by using the pronoun, he must draw from information outside of the pronoun itself. We have found that by focusing on the ubiquitous task of **reference**, large portions of context can be constrained so that others can be investigated.

In general, it is the hearer's job to use information from the context to determine which of a presumed set of meanings the speaker intended when she chose to use a particular *lexical item* (e.g., a morpheme, word, or idiom). And although the lexical items in an utterance arrive in a linear order, the hearer may need to backtrack and revise his decisions based on subsequent information. If the hearer does not revise an incorrect decision, he will likely fail to arrive at the speaker's intended meaning. The linguistic phenomena of ambiguity and vagueness are

---

two root causes of interpretive decision points. In certain contexts, these phenomena make linguistic communication efficient for humans [3, 4] and extremely challenging for computational models.

Our goal is to build a computational model of reference that is able to represent only the relevant linguistic choices, and then make the correct decisions. We approach the problem in two stages: (1) making the system *expressive* enough to capture desired linguistic phenomena by ensuring the system is capable of representing all choice points that lead to the desired output[1] and (2) finding a control algorithm that minimizes the number of choice points considered to produce the desired output.

Unfortunately, theories of discourse context are rarely defined precisely; and so it is difficult to separate the components of context that influence lexical items from those that do not. We attempt to rectify this: in section 1.1, we summarize several ways context can influence the interpretation of referring expressions, and in section 1.3, describe a constrained communication task in which context's influence on lexical items can be modeled directly. Afterward, we present an incremental model of reference interpretation that defines the meanings of vague (gradable) and lexically ambiguous adjectives using information from the on-line denotation.

1.1   What components of discourse context influence reference interpretation?

In the study of language, the term "context" has been used to connote a wide range of information that is available to the speaker or hearer. A skeptic might take it to mean *any information that outside the scope of the theory at hand.*

Tomasello described **discourse context** as "information that is available to [both speaker and hearer] in the environment, along with what is 'relevant' to the social interaction, that is, what each participant sees as relevant and knows that the other sees as relevant as well—and knows that the other knows this as well, and so on, potentially ad infinitum. This kind of shared intersubjective context is what we may call following [5] **common ground**...it takes [hearer and speaker] beyond their own egocentric perspective on things" [6, pp. 76].

Of course, common ground is a fiction: in addition to being paradoxically recursive, neither speaker nor hearer are omniscient so neither could ever know the true common ground. However, as a theoretical concept it may still be useful to envisage such an idealized state both the speaker and hearer's inferential processes work toward in order to make the reference task succeed. As such, the speaker and hearer each have their own notions of success. The speaker wants to convey the referents to the hearer, so she must take into account what he knows or is capable of inferring. Similarly for the hearer, the speaker's act of reference contains "an implicit assurance that he has enough information to uniquely identify the referent, taking into account the semantic content of the

---

[1] Because our system is *incremental*, the "desired output" of a referring expression can be evaluated at intermediate stages.

referring expression and information from the context, whether situational (i.e. currently perceivable), linguistic, or mental (i.e. memory and knowledge)" [7].

For reference tasks, the knowledge speaker and hearer can be expected to have minimally includes:

Task. The speaker and hearer's shared tasks determine what is relevant and important to them, and thus their communication goals as well. Using the pragmatic theory of [8], the information needs of the task constitute its **questions under discussion**, which are a central impetus of communication. From our computational perspective, we take a *referential* question under discussion to be an *unbounded typed variable* in a plan. The question is answered when the variable is bound to a knowledge representation that meets certain type restrictions. Questions under discussion give rise to communication goals, which are fulfilled by communication acts toward these goals (e.g., speaking, gesturing). For reference tasks, the communication goal is at least in part referential: to make the intended referent(s) mutually known to hearer and speaker (i.e., in the common ground).

Referential domain. Entities in the **environment**, which are mutually perceived, along with concepts from **background knowledge** constitute potential targets of referring expressions.

Dialogue history. The speaker and hearer can be expected to remember the previous dialogue acts. For reference, this is especially important because after a speaker introduces a referent to discourse, she typically mentions it again—often using abbreviated referring expressions [9–11]. From a computational perspective, the referents in the dialogue history could be thought of as *symbol table* used by compilers and interpreters to map each symbol to its type, scope and value—namely, its location in memory.

Instead of describing these contextual constraints individually, we will introduce an abstraction called the **context set**. It is a construct from theoretical linguistics that represents the "live options"—viable candidates for an interpretation process, which evolves over the course of dialogue [12]. It is the hypothesis space of interpretations. For utterances outside of reference, the concept of "what constitutes an interpretation" is difficult to pin down; however, for reference tasks, the context set can be seen as the **referential domain** plus all of its combinatoric possibilities.

1.2   Characterizing the two reference tasks

The referential domain and its valid means of combination are both constrained by information from the task, dialogue history, and lexical-semantic knowledge. By constraining each of these elements, then, at least for reference tasks, we can replace the illusive concept of context with a single construct, the context set, which expresses the sum of all contextual constraints on the targets for interpretation. We will now attempt to formalize the broader communication tasks of the speaker and hearer. When referring, the speaker and hearer complete

two structurally similar tasks. The speaker completes a **referring expression generation (REG)** task: given an initial *context set* (defined in Section 1.4) and a designated member of it called the *target set*, she produces a *referring expression* which she expects will enable the hearer to infer her intended target set from the rest of the elements in the context set, called *distractors* [13]:

$$\mathrm{REG}(\mathsf{context\ set}, \mathsf{target\ set}) \rightarrow \mathsf{referring\ expression} \tag{1}$$

A hearer completes a **referring expression interpretation (REI)** task: given a referring expression, his goal is to jointly infer the context set and the targets that the speaker intended:

$$\mathrm{REI}(\mathsf{referring\ expression}) \rightarrow \langle \mathsf{context\ set}, \mathsf{target\ set} \rangle_1 \dots \langle \mathsf{context\ set}, \mathsf{target\ set} \rangle_n \tag{2}$$

Reference tasks do not always succeed. We define a **reference failure** as a mismatch between the speaker's intended target set and the one (or ones) yielded by the hearer's interpretation. If the referring expression leads the hearer to generate multiple plausible interpretations (e.g., $n > 1$), we call such referring expressions *uncertain*. In the next section, we describe some of the issues that lead to these uncertain referring expressions, which in turn commonly lead to reference failures.

1.3   Restricting the discourse context

Although all of the aforementioned components of context can potentially impact reference interpretation [14, 11, 15], to avoid their influence we can restrict the task setting so that many aspects of context *prior* to an utterance are constrained. This allows us to investigate and model how words interact with context. We do this by:

– using a **referential domain** that is co-present [16], which we achieved by using *visual scenes*. Most of the referring expressions described in this paper will be interpreted with respect to one of two referential domains, CIRCLES and KINDLES, which are expressed as co-present visual scenes and therefore assumed to be in the speaker and hearer's common ground:

Fig. 1. The Amazon KINDLE referential domain containing 5 referents: k_1, k_2, k_3, k_4 and k_5.

Fig. 2. The CIRCLE referential domain containing referents: c_1, c_2 and c_3.

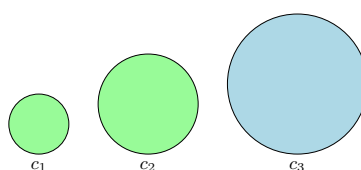- when using descriptive referring expression (i.e., noun phrases), ensuring that it is purely referential and does not serve ulterior communication goals
- embedding referring expressions within a consistent and simple task context. Presumably there is always an implicit task context that is motivating the hearer to cooperate, so it is important that at least the local goal be controlled [17, pp. 313]. In psycholinguistic research about reference production, it is routine to give hearer's an imperative sentence requiring them to manipulate physical referents (e.g., *Pass me the green cup*), this provides a neutral task context that is consistent across subjects.
- assuming there is no dialogue history, and we have "one-shot referring expressions" [9])

Such one-shot, task-neutral, purely referential referring expressions in co-present visual domains are the focus of the paper. They allow us to investigate out the context set evolves over the course of a single interpretation; and allow us to characterize the discourse context exclusively in terms of the elements in a referential domain (as we do in Section 1.4).

1.4   Formalizing the context set: the hypothesis space of interpretations

By restricting our attention to one-shot, purely referential referring expressions, we can represent the context set in explicit detail.

A candidate **interpretation**, a member of the context set, can be viewed *intensionally* as some epistemic data structure or *extensionally* by its **denotation**, denotation($\cdot$),[2] which enumerates the groups of targets that match its intensional description. For example, given the CIRCLES reference domain and the referring expression "the biggest one", the denotation is all of the possible targets that are consistent with its intension (which we represent using belief states): denotation("the biggest one") = $\{c_3\}$. Although the semantic, intensional meaning representations for "the biggest one" and "the blue circle" are presumably very different, their denotations are the same: denotation("the biggest one") = denotation("the blue circle"). The denotation of a referring expression provides one means to probe how people represent linguistic meanings—e.g., it is easy enough to ask a subject to "*select* the second biggest green one" from a visual scene.

Under our constrained reference task, we can describe the upper bounds on *denotational complexity* for context sets entirely in terms of the referential domain, R, for models with different degrees of expressiveness. Keep in mind that there may be a large number of varied intensional descriptions for each denotation, whose number depends on the particular representation used (e.g., typically a logical form or attribute-valued matrix). The context set for a system that interprets (or generates) referring expressions that can only refer to *single targets* contains $|R|$ distinct denotations. For example, the singleton context set for the CIRCLES referential domain contains $|R| = 3$ elements, and would only be able to deal with referring expressions that refer to $\{c_1\}$, $\{c_2\}$ or $\{c_3\}$. Examples: denotation("the blue circle") = $\{c_3\}$ and denotation("the red triangle") = $\emptyset$.

If an interpreter can refer to sets, as do the approaches described in [18–20], it explores a hypothesis space containing $2^{|R|} - 1$ denotations, which is analogous to a **belief state**[3] about the singleton domain, R. An example in the CIRCLES domain: denotation("the green circles") = $\{c_1, c_2\}$.

To handle all of the linguistic phenomena we will describe in Section 2, we will want to represent multiple interpretations about sets (due to unspecific descriptions, vagueness and ambiguity) so our hypothesis space contains $2^{2^{|R|}-1} - 1$ interpretations. This permits us to represent, for example, the lexical ambiguity of the word "biggest", which in the KINDLES domain, has two senses: it can refer to the Kindle with the biggest screen ('big$_1$'), or the one with the biggest hard drive ('big$_2$'): denotation("the biggest kindle") = $\{k_4\} \oplus \{k_5\}$. We treat these two interpretations as mutually exclusive, represented by the $\oplus$ symbol. This state-space grows large quickly, however: for the CIRCLE domain, where $|R| = 3$, there are 127 denotations; while for KINDLE, where $|R| = 5$, there are over two billion. Fortunately, there are many ways to avoid this complexity [21].

---

[2] Commonly represented by $[\![\cdot]\!]$.

[3] In artificial intelligence, the power-set of a set of propositions is commonly used to represent a *belief state*, which characterizes incomplete knowledge about an underlying set. Beliefs are an abstraction of any lower layer, so beliefs can be about beliefs about beliefs. Here, we use second-order belief states that are two layers removed from the referential domain to represent uncertainty about representations of sets of candidate targets.

2 Context-sensitive Referring Expressions

2.1 Lexical ambiguity

Lexical ambiguity is when a *lexical unit* maps to multiple meanings. Determining whether a lexical unit's meanings (*senses*) are truly the same can be challenging, because it is common for the various senses to all be related somehow (e.g., with polysemous words) [22, 23]. In other cases, a lexical unit's different senses are clearly disjoint (e.g., homonyms), as with the canonically ambiguous noun "bank", whose meanings include $bank_1$, *a financial institution*, and $bank_2$, *the land border along of a river.* Upon encountering ambiguity, the reader is confronted with a choice between alternative meanings.

(1.a) Let's go stop by the *bank*
(2.a) Let's go fish by the *bank*

Lexical ambiguity is also constrained by the context set. If (1.a) were uttered in a rural community that did not have any financial institutions but did have accessible rivers, we would expect the meaning of "bank" to only describe $bank_2$ (river); and, because it is singular and definite (i.e., begins with 'the'), it presupposes that the referring expression along with the context set is enough for the hearer to arrive at a target set containing a single river bank.

Psycholinguistic studies have given compelling evidence that readers' disambiguation choices are influenced by the referents that are available in the context set, and that readers update their context sets incrementally and frequently while reading. Such research, surveyed in [17], present subjects (or hearers) with a spoken referring expression in a visual scene (ensuring co-presence and common ground), and monitor the subject's eye movements (which tend to focus on the working target interpretation). For compositional theories, this implies that humans incrementally *evaluate* the semantic representation to yield its denotation [24], which implies syntax and semantics are tightly coupled. What this means is that when interpreting the referring expression online, by the time we have arrived at the ambiguous choice point, the context set has already updated to reflect the partial information imposed by the verb's selectional constraints and may bias us to favor one particular resolution:

(1.b) Let's go stop by the *bank* $\rightarrow$ $bank_1$ (finance)
(2.b) Let's go fish by the *bank* $\rightarrow$ $bank_2$ (river)

For our purpose of modeling human performance in the REG and REI tasks, these issues raise important computational questions, including:

Q1 What kind of interpreter will incrementally perform syntactic and semantic analysis, and allow the denotation to be available at each choice point?
Q2 At what granularity should choice points be represented?
Q3 Are the multiple senses of an ambiguous lexical unit first generated and subsequently filtered when they are incompatible with the context set, or generated as a function that operates on the context set?

2.2 Garden paths and incremental interpretations

To illustrate the highly incremental nature of interpretation, observe that the reader's initial disambiguation choices, which produced the interpretations alluded to in (1.b) and (2.b), can be reversed by adding linguistic context. These examples produce the so-called *garden-path effect*, because they cause the reader to revise his initial ambiguity resolution decision in the face of new conflicting evidence:

(1.c) Let's go stop by <u>the *bank* of the Charles River</u> $\rightarrow$ bank$_2$ (river)
(2.c) Let's go fish by <u>the *Bank* of Commerce</u> $\rightarrow$ bank$_1$ (finance)

A reader, despite the fact the text of (1.a/1.b) is a prefix of (1.c), will react to the additional context flexibly by switching from one sense of *bank* to another. This means the semantic content of her interpretation is *non-monotonic*: the combined meaning up to word w_{i+1} may not have been included in (or entailed from) the meaning up to word w_i; and this implies for the usage of "bank" in this instance that "not all mouths of rivers are financial institutions."

Q4 Are multiple interpretations constructed simultaneously (e.g., by taking all choices in parallel), or just one at a time (e.g., maintaining a single best interpretation and then backtracking when necessary)?

2.3 Vagueness and gradable adjectives

Another threat to recovering the speakers' intended reference is **vagueness**. The term "vagueness" itself is lexically ambiguous. Linguists and laypeople typically use it as vagueness$_1$ (insufficient information), which means (autologically), *insufficiently informative for the current purposes* [25]. An example of vagueness$_1$ (insufficient information) is:

(3) Let's meet for dinner at <u>a restaurant</u>

when there are more than one restaurants in the referential domain. The other sense, vagueness$_2$ (borderline cases), is better known to philosophers of language, and connotes something more specific: *predicates with unclear denotations—i.e. denotations containing borderline cases* [26]. This is a symptom found in many *gradable (scalar) adjectives* such as 'tall,' 'big,' and 'short.'

(4) Let's watch <u>a *short* movie</u> $\rightarrow$ *short for movie: less than 2 hours? 1.5 hours?*

Gradable adjectives impose a *relational constraint* between ordered values of an attribute that varies between referents in the target set and its distractors [27]. They can be problematic when there are referents whose values for that attribute are in the middle of the ordering. For example "<u>expensive restaurants</u>" may definitely include restaurants whose average meal costs $10 or less, definitely exclude those whose average meal price is more than $40, but not lead to any consensus for restaurants who have an average price in between.

Referents that are "in between" are *borderline cases*, and in referring expressions these can cause reference failures [4]. For you to succeed at interpreting (4), "a short movie", you must pick (a) a *comparison class* that defines the set of movies relevant to your comparison, and (b) a *standard of comparison* that delineates $\text{SHORT}(x_i)$ from $\neg\text{SHORT}(x_{i-1})$ for the ordered elements x in the comparison class. Under the constrained reference task we described in 1.4, the comparison class (a) can be assumed to be the referents that are consistent with the current interpretation, leaving the standard, (b), as our main concern. Depending on how you set this standard, the interpreter may arrive at different interpretations.

Q5 What representation allows interpretations to be individuated in a way that distinguishes vagueness$_1$ (insufficient information), vagueness$_2$ (borderline cases), and ambiguity?

3 Our Approach: Planning in Belief Space

We have developed a fast belief-state planner, AIGRE, [4] that can generate and interpret simple English referring expressions. In order to represent the combined constraints of all information in the context set up to the decision point of an ambiguous or vague lexical unit (i.e., **Q1**), AIGRE avoids traditional pipeline architectures and takes an integrated "lexicalized approach," following [28–31], in which each surface form (lexical unit) and its syntactic, semantic, and (conventional) pragmatic contributions are collectively represented in a *lexical entry* (**Q2**) and come into effect at the same time. This allows us to interleave decisions about *what* to say and *how* to say them, and express them in a uniform manner [32].

This formulation reduces the entire generation task (REG) to choosing the actions whose effects achieve the speaker's communicational goal and putting them in the correct order, a formulation that bears a strong analogy with **automated planning**. However, instead of changing the state of the world, the actions change *belief-states*, which represent complete interpretations—implicitly representing all possible targets. For REI, we complete a **plan recognition** task of a similarly searching for a plan. Rather than being directed by a communication goal, fore the REI task, the action sequence is constrained to those that can produce the observed utterance and by a language model that enforces syntax constraints. The belief-state planner using heuristic search to find a sequence of lexical units that map the initial belief state onto a target belief state (see [21] for more details).

Because belief states are complete interpretations, the planner is incremental and the denotations for all candidate interpretation can be output at any stage. The initial belief state is one of complete uncertainty—all $2^{|R|}$ target sets are possible; only as it accumulates information do the possibilities decrease. As we

---

[4] **A**utomatic **i**nterpretation and **g**eneration of **r**eferring **e**xpressions. In French, it means "sour".

mentioned in Section 1.4, belief states have a strong connection with *context sets*, which we take to represent the combined meaning of all belief states at any point of the interpretation process.

3.1   Representing context-sensitive actions

AIGRE's lexicon is comprised of lexical units, which are belief-changing actions. Currently, AIGRE's lexicon is restricted to nouns and prenominal modifiers. Each action/word is an instantiation of an action class and has (1) a syntactic category (part of speech), (2) a lexical unit, (3) a specific semantic contribution—determined in part by its syntactic category, (4) a fixed lexical cost, and (5) a computed effect cost. Actions are defined by instantiating class instances, for example:

```
– GradableAdjective(lexeme='big', attribute='size')
– CrispAdjective(lexeme='blue', attribute='color', value='blue')
```

When instantiating an action, the first argument is its lexeme in its *root form*; the class' initialization method uses the root lexeme to also instantiate variant actions for each derivative lexical unit (e.g. plural, comparative, superlative, etc). Syntax constraints are expressed in the transitions between states by the *action proposal function* which is responsible for filtering the actions depending on the current state.

The actions, which generate lexical items, operate on an interpretation and yield successors. Ambiguity and gradable meanings are modeled using non-deterministic actions (see Algorithms 1 and 2): they receive a belief state as input and lazily generate 0 or successors,[5] depending on the contents of the belief state. Thus part of a word's meaning can come from its interaction with the belief state. Not having any effects is analogous to not having its preconditions satisfied. This lends itself to a procedural semantics where the meaning of a given word can interact directly with the contents of a given hypothesis (interpretation, belief state) in the context set (**Q3**).

3.2   Crisp and Graded Adjectives

The semantic contribution of a CRISPADJECTIVE (e.g. *silk, John's, prime, pregnant*), is akin to traditional assignment: an attribute in the intensional representation is assigned a specific value. It is only non-deterministic when the parent state has members with multiple properties with the same attribute name.

Gradable adjectives do not require a `value` parameter when they are initialized, because their value comes from the context set they are applied to. Like CRISPADJECTIVE, a GRADABLEADJECTIVE also iterates through each sense of the attribute, but adds *intervals* values to the belief state's description rather than atomic values:

---

[5] Rather, they generate *effect functions* that operate on states. To simplify the exposition, we pretend our actions are like those of typical planners and generate successor states.

Data: Initialized with an attribute name, a, and value, v
Input   : A search node, S, containing a belief state
Output: A successor search node, $\hat{S}$

foreach a *in S.getAttributesByName(attributeName)* **do** breadth-first iteration of referents' attributes $= a$

> create copy of S, named $\hat{S}$;
>
> attempt to merge $\hat{S}.a$ with value v;
>
> yield $\hat{S}$

end

Algorithm 1: Effects of a CRISPADJECTIVE. Belief states' properties represent partial information using *ranges* of values; they are updated by `merge` rather than `setter` methods (based on the *cell* datastructure of [33]).

Data: Initialized attribute, a, and $minimize = False$
Input   : A search node, S, containing a belief state
Output: Yields 0 or more search nodes, $\hat{S}$

foreach a *in S.getAttributesByName(attributeName)* **do** breadth-first iteration of referents' attributes $= a$

> **foreach** v *in S.getUniqueValues(a,minimize)* **do**   iterate context set's unique, sorted values of a
>
> > create copy of S, named $\hat{S}$;
> >
> > **if** *minimize* **then**
> >
> > > attempt to merge $\hat{S}.a$ with interval $(-\infty, v]$
> >
> > **else**
> >
> > > attempt to merge $\hat{S}.a$ with interval $[v, \infty)$
> >
> > **end**
> >
> > yield $\hat{S}$;
>
> **end**

end

Algorithm 3: The lexical entry for a *gradable adjective* action: its job is to lazily yield successors for each sense (ambiguity) for each value (vagueness). Actions act on *belief states*, datastructures that represent an *intensional* description of the context set.

Gradable adjectives yield an effect for each same-named attribute (ambiguity) for each value (vagueness$_2$ (borderline cases)). For example, given the action BIG$_{JJ}$ applied to b_0 (about the KINDLE referential domain) yields a separate effect for each unique value of each unique attribute-path named size, producing 6 effects in total, BIG$(b_0) \rightarrow e_0, e_1 \ldots e_6$. Three of the effects, when executed, would create target.size properties and the other three create target.hard_drive.size properties. Gradable adjectives' values are represented with an interval. For "big," e_0 would add (if it doesn't already exist) and then attempt to merge the belief state's target.size value with an interval beginning at the largest size value ($\theta = 7$) of a referent consistent with b_0.

3.3   Interpreting "the big ones" in two domains

To illustrate how AIGRE interprets a referring expression, "the big ones", which contains both ambiguity and vagueness$_2$ (borderline cases), we interpret it with respect to both domains.
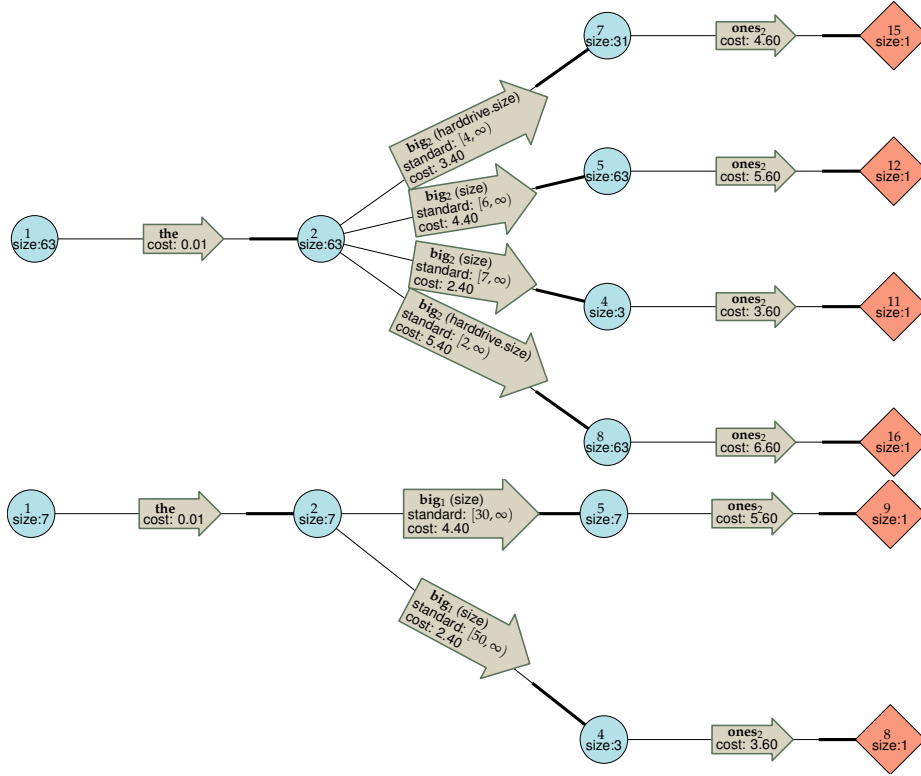

Fig. 3. A side-by-side comparison of the top-to-bottom **search graphs** for interpreting "the big ones" in both the KINDLES (top) and SHAPES (bottom) domains. Each blue node represents a belief state, the red diamonds are *goal* belief states, and the edges correspond to *choice* about which (sense of a) lexical unit to apply.

For the KINDLES domain, denotation("the big ones") = $\{k_4, k_5\} \oplus \{k_2, k_3, k_4, k_5\} \oplus \{k_1, k_2, k_3, k_4, k_5\}$. For CIRCLES, denotation("the big ones") = $\{c_2, c_3\} \oplus \{c_1, c_2, c_3\}$. These interpretations correspond to the combined choices from the lexical ambiguity[6] ('big$_1$' versus 'big$_2$'), the various ways one can set the standard of comparison for each sense of 'big'—at one extreme value for the standard, only the biggest element is BIG and at the other extreme *all* elements are considered

---

[6] The lexical ambiguity is only applicable in the KINDLES domain, because each referent has two size properties, whereas in CIRCLES they only have one.

BIG, along with other linguistic constraints like number agreement (e.g. plural requires more than one element in the context set).
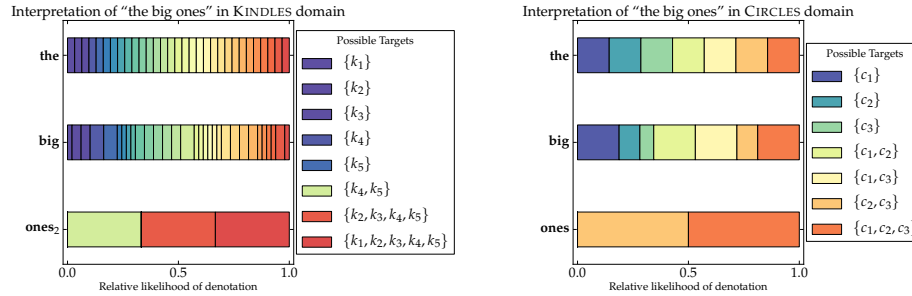

Fig. 4. An alternative view of the same two interpretations, showing the evolution of the context set: the *combined* denotations from each alternative interpretation. Each row corresponds to a column of states in the search graphs in Figure 3, excluding the initial belief state (node 1). For the first row in KINDLES, "the," is node 2 and the last row, "ones," is the denotations of interpretations/belief states of nodes 7, 8, 9 and 10 combined. The size of the member indicates its relative likelihood, determined by the inverted sum of its interpretation's relative costs.

4   Related Work

For REG with ambiguity, [34] described an approach to learning how to predict whether a given referring expression contains a structural ambiguity. For work on vagueness, [18, 35]'s system, VAGUE, generated referring expressions that included gradable adjectives, but managed to do so in a deterministic way. The authors intentionally avoided plural gradable adjectives in their base form because of the arbitrariness of their meanings, so they were required to produce "the two biggest ones" rather than "the biggest ones" when describing multiple items. With AIGRE, we embrace the non-determinism and control the search such that most common standards of comparison are chosen first, while the less common standards remain possibilities through the use of backtracking. In addition, there are different characterizations of gradable adjectives' semantics: [36] uses probabilities and the fuzzy-logic community does so using gradual membership functions [37, 38].

Finally, we would like to note that [39] presented a different formalization of the context set, in which its elements also contained weights that indicated its *salience*. This view is not incompatible with ours; salience can be viewed as *prior distribution* over the entities in the context set. In our visual scene paradigm, we make a simplifying assumption that all of the groups of referents are equally salient.

5   Algorithmic Evaluation: The scalability of the lexicon

A key bottleneck to scalability is the *number of relevant actions* that need to be considered during generation and interpretation. For interpretation, the actions were highly constrained to those that appeared in the referring expression, so search space was constrained enough that it could be generated and explored in its entirety within less than a second. However for generation, the worse case branching factor was the total number of actions. Worse, in our representation the effects of vague and ambiguous actions proliferate: if the adjective BIG has s senses, and there are r referents compatible with the belief state, then it can yield as many as $s \times r$ successors. To avoid this complexity, ambiguous words' *senses* and vague words' *standards* are only generated when needed (**Q4**).

We compared three methods to control search: (1) *A* search*, an optimal strategy that picks the node with the lowest combined cost and estimated distance, and expands all of its successors (2) *best-first search*, which picks the node with the lowest estimated distance and expands all of its successors, and (3) *stochastic hill climbing* (or stochastic local search), which picks the node with the lowest estimated distance and then (sometimes) stops expanding it when it generates a better successor. As you can see in Figure 4, the stochastic hill-climbing time scaled linearly while the others scaled exponentially.
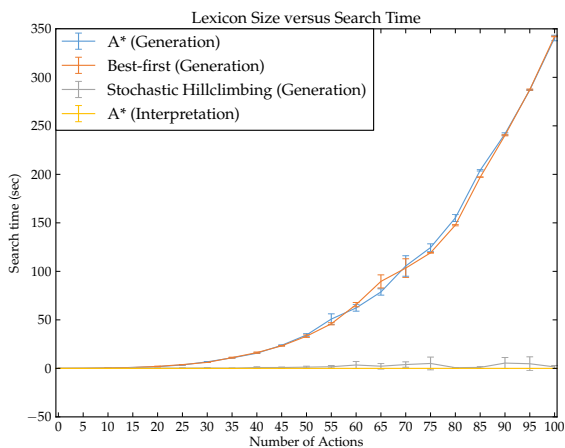

Fig. 5. Each condition was run for 5 trials.

6   Conclusion

Viewing reference generation and interpretation as a search through belief states forces one to see the problems through procedural lenses. From this perspective, it was clear that the numerous shades of meaning introduced by ambiguous and vague modifiers should be confined to those senses that are *relevant*. We achieved by defining the possible meanings of ambiguous and vague modifiers using properties of the targets in the partial interpretation. With backtracking, we avoided generating every possibly-relevant meaning at once by structuring the search space to *make the common senses of lexical units easiest to find, while enabling the less common meanings to still be possible.*

For vagueness$_2$ (borderline cases) this seems straightforward: start at the most conservative meaning (i.e., 'big' = 'biggest') and work backward when needed; however, the ordering is not obvious for lexical ambiguity. Our next steps will be to learn these orderings from human behavioral data.

References

1. Wilson, D., Sperber, D.: Meaning and Relevance. Cambridge University Press (2012)
2. Reiter, E., Dale, R.: Building natural language generation systems. Cambridge University Press, New York, NY, USA (2000)
3. Piantadosi, S.T., Tily, H., Gibson, E.: The communicative function of ambiguity in language. Cognition (2012)
4. Van Deemter, K.: Not exactly: In praise of vagueness. OUP Oxford (2010)
5. Clark, H.H.: Using language. (1996)
6. Tomasello, M.: Origins of Human Communication. MIT Press (2008)
7. Cruse, A.: Meaning in Language: An Introduction to Semantics and Pragmatics. Oxford University Press (2010)
8. Roberts, C.: Context in dynamic interpretation. The handbook of pragmatics (2004)
9. Brennan, S.E., Clark, H.H.: Conceptual pacts and lexical choice in conversation. The Journal of Experimental Psychology (1996)
10. Branigan, H.P., Pickering, M.J., Pearson, J., McLean, J.F.: Linguistic alignment between people and computers. Journal of pragmatics (2010)
11. Goudbeek, M., Krahmer, E.: Alignment in Interactive Reference Production: Content Planning, Modifier Ordering, and Referential Overspecification. Topics in Cognitive Science (2012)
12. Stalnaker, R.C.: Assertion Revisited: On the Interpretation of Two-Dimensional Modal Semantics. Philosophical Studies (2004)
13. Dale, R., Reiter, E.: Computational interpretations of the Gricean maxims in the generation of referring expressions. Cognitive Science (1995)
14. Clark, H.H., Schreuder, R., Buttrick, S.: Common ground at the understanding of demonstrative reference. Journal of Verbal Learning and Verbal Behavior (1983)
15. Janarthanam, S., Lemon, O.: Learning lexical alignment policies for generating referring expressions for spoken dialogue systems. In: Proceedings of the 12th European Workshop on Natural Language Generation. (2009)
16. Clark, H.H., Marshall, C.R.: Definite reference and mutual knowledge. Psycholinguistics: critical concepts in psychology (2002)
17. Tanenhaus, M.K.: Spoken language comprehension: Insights from eye movements. Oxford handbook of psycholinguistics (2007)
18. van Deemter, K.: Generating vague descriptions. In: INLG '00: Proceedings of the first international conference on Natural language generation. (2000)
19. Stone, M.: On identifying sets. In: INLG '00: Proceedings of the first international conference on Natural language generation. (2000)
20. Horacek, H.: On Referring to Sets of Objects Naturally. In: Natural Language Generation. (2004)
21. Smith, D., Lieberman, H.: Generating and Interpreting Referring Expressions as Belief State Planning and Plan Recognition. Proceedings of the 14th European Workshop on Natural Language Generation (2013)
22. Palmer, M.: Consistent criteria for sense distinctions. Computers and the Humanities (2000)
23. Allen, J.F.: Word Senses, Semantic Roles and Entailment. 5th International Conference on Generative Approaches to the Lexicon
24. Altmann, G., Steedman, M.: Interaction with context during human sentence processing. Cognition (1988)

25. Dubois, D.: Have fuzzy sets anything to do with vagueness. This volume (2011)
26. Graff, D.: Shifting sands: An interest-relative theory of vagueness. Philosophical Topics (2002)
27. Sedivy, J.C., K Tanenhaus, M., Chambers, C.G., Carlson, G.N.: Achieving incremental semantic interpretation through contextual representation. Cognition (1999)
28. Stone, M., Doran, C., Webber, B.L., Bleam, T., Palmer, M.: Microplanning with communicative intentions: The spud system. Computational Intelligence **19**(4) (2003) 311–381
29. Koller, A., Stone, M.: Sentence generation as a planning problem. Annual Meeting of the Association of Computational Linguistics (2007)
30. Bauer, D.: Statistical natural language generation as planning. Master's thesis, Department of Computational Linguistics, Saarland University, Saarbrücken, Germany (2009)
31. Koller, A., Gargett, A., Garoufi, K.: A scalable model of planning perlocutionary acts. In: Proceedings of the 14th Workshop on the Semantics and Pragmatics of Dialogue. (2010)
32. Appelt, D.E.: Planning English referring expressions. Artificial Intelligence (1985)
33. Radul, A., Sussman, G.: The art of the propagator. Technical report, Technical Report MIT-CSAIL-TR-2009-002, MIT Computer Science and Artificial Intelligence Laboratory (2009)
34. Khan, I., van Deemter, K., Ritchie, G.: Managing ambiguity in reference generation: the role of surface structure. Topics in Cognitive Science (2012)
35. van Deemter, K.: Generating Referring Expressions that Involve Gradable Properties. Computational Linguistics (2006)
36. Lassiter, D.: Vagueness as probabilistic linguistic knowledge. Vagueness in Communication (2011)
37. Zadeh, L.A.: Fuzzy logic and approximate reasoning. Synthese (1975)
38. Hersh, H.M., Caramazza, A.: A Fuzzy Set Approach to Modifiers and Vagueness in Natural Language. (1976)
39. Krahmer, E., Theune, M.: Context sensitive generation of descriptions. (1998)