Givenness Hierarchy Theoretic Referential Choice in Situated Contexts

Poulomi Pal, Grace Clark and Tom Williams
{poulomipal,geclark,twilliams}@mines.edu
MIRRORLab, Colorado School of Mines
1600 Illinois Street, Golden, CO 80401 USA

Abstract

We present a computational cognitive model of referential choice that models and explains the choice between a wide variety of referring forms using a small set of features important to situated contexts. By combining explainable machine learning techniques, data collected in situated contexts, and recent computational models of cognitive status, we produce an accurate and explainable model of referential choice that provides an intuitive pragmatic account of this process in humans, and an intuitive method for computationally enabling this capability in robots and other autonomous agents.

Keywords: cognitive status; anaphora generation; natural language generation; referential choice

Introduction

A central feature of human language is the use of reference, through which interlocutors pick out or introduce entities. When making reference to target referents, people use a wide variety of referring forms, including definite and indefinite noun phrases (i.e., *the* ⟨*N'*⟩, *a* ⟨*N'*⟩) and various deictic and pronominal forms (e.g., *it, this, that, this* ⟨*N'*⟩, *that* ⟨*N'*⟩)[1]. A vast amount of research has been performed seeking to explain and model how humans generate definite descriptions, as described by Van Deemter (2016); and to enable robots and other autonomous systems to efficiently generate effective and natural referring expressions, as described by Gatt and Krahmer (2018). However, there has been substantially less work in both communities seeking to explain and model the use of referring forms beyond definite descriptions.

As highlighted by Gundel, Hedberg, and Zacharski (1993), humans use this wider variety of referring forms in curiously flexible and strategic ways. Referring forms such as *it, this*, and *that*, for example, provide little semantic information about their targets, and yet humans strategically use these forms not only to express themselves more concisely but moreover to provide subtle cues that allow their interlocutors to more quickly and effectively identify target referents. Moreover, referential choice (the selection between these different forms) is widely accepted to be an important first step in the natural language process (Kibrik et al., 2016), made before descriptive content is considered for inclusion (Krahmer & Van Deemter, 2012). Accordingly, understanding and modeling the process by which humans select between referring forms is critical both from a psycholinguistics perspective as well as for those in the artificial intelligence community seeking to build robots and other autonomous agents capable of effectively and naturally communicating with human collaborators through natural language.

Various theories of reference have sought to explain the types of factors that determine what referring forms can be used felicitously in different contexts. For example, and of particular interest in this work, the *Givenness Hierarchy* (Gundel et al., 1993) suggests that different referring forms signal different *cognitive statuses*, with "this" signaling that the speaker believes their referent to be at least *Activated* in the current conversation and/or in the mind of their interlocutor (Rosa & Arnold, 2011). But while theories such as the Givenness Hierarchy provide critical linguistic insights, they do not attempt to explain or algorithmically model the cognitive *process* of referential choice. And while there has been vast literature from both psycholinguistic and artificial intelligence perspectives on both the understanding of referring language (including deictic and anaphoric language), and the generation of non-anaphoric language, there has been relatively little work on computationally modeling the selection between anaphoric forms during the natural language generation process. Finally, work on computational modeling of referential choice has typically been focused on predicting whether or not to use a pronoun using the set of features that can be carefully annotated in textual domains.

We propose a cognitive model of referential choice that predicts the choice between a wide variety of referring forms using a small set of features well established as important in situated contexts. By combining explainable machine learning techniques, data collected in situated contexts, and recent work on the computational modeling of Givenness Hierarchy theoretic notions of cognitive status, we produce an accurate and explainable model of human referring form that provides an intuitive account of this process in humans and an intuitive method for computationally enabling this capability in robots and other autonomous agents.

[1]The examples used in this paper are in English, but the variety of referring forms we consider and the way they are used has been well documented across a wide variety of languages, including Mandarin, Japanese, Spanish, Russian, Eegimaa, Kumyk, Ojibwe, Arabic, Irish, Norwegian, Persian, and Turkish (Hedberg, 2013).

Related Work

Linguistic Models of Referential Choice

As discussed by Arnold and Zerkle (2019), linguistic models seeking to explain the use of different referring forms, especially pronouns, fall into two broad categories. *Rational* model seek to explain the production of pronouns as a matter of egocentric analysis of the costs of generating and processing different referring forms, with pronouns typically preferred due to their ease of production (Aylett & Turk, 2004; Frank & Goodman, 2012). *Pragmatic* models suggest that reduced forms like pronouns are special forms that humans have learned to associate with referents that are highly activated or focused; an allocentric framing that highlights pronouns' ability to enhance not just ease of speaker production but also ease of listener interpretation. As argued by Arnold and Zerkle (2019), while pragmatic models may differ in terms of the theoretical constructs they use to explain differences in information status (i.e., focus (Grosz, Joshi, & Weinstein, 1995; Brennan, Friedman, & Pollard, 1987; Grosz & Sidner, 1986), salience and accessibility(Ariel, 1991), or givenness (Gundel et al., 1993), and accordingly differ as to whether they regard information status as a difference of linguistic category, psychological state, or presumption of psychological state, all of these models share a common assumption that use of different referring forms is grounded in a relationship between discourse status and psychological states, especially memory and attention.

While both classes of theories have promise, neither does a terribly good job at explaining the choice between different *types* of reduced expressions, e.g., between *it, this,* and *that*. As Arnold describes, this problem is particularly salient for rational models, which should suggest much more frequent use of reduced forms than is actually seen in practice, and fails to account for patterns of selection between referring forms that are equally short. Moreover, Arnold and Zerkle highlights that linguists have tended to focus on explaining individual phenomena, and that there has not to date been an attempt to provide a comprehensive explanation for reference production as a whole. This perspective aligns with that argued by Grüning and Kibrik (2005) who highlight that many linguists have focused narrowly on individual factors that may impact referential choice, such as linear (linguistic) distance (Givón, 1983), rhetorical distance (Fox, 1993; Mann, Matthiessen, & Thompson, 1989), and narrative episodic structure (Tomlin, 1987; Marslen-Wilson, Levy, & Tyler, 1982). Finally, we would add to this analysis that models of both varieties suffer from an overemphasis on textual analysis, failing to adequately model the aspects of situated contexts.

In this work, our first goal is thus to provide a more comprehensive account of referential choice, that accounts for generation of a wide set of referring forms in situated contexts. Specifically, we take a cognitivist perspective grounded in the theory of the *Givenness Hierarchy* (GH) (Gundel et al., 1993). The Givenness Hierarchy suggests that a speaker's choice of referring form is made based on their assumptions as to the cognitive status of their target within the mind of their interlocutor or within the conversation. If they assume their target is "In Focus" they may choose a form such as *it*; if they assume their target is "Activated" (a category which subsumes "In Focus") they may choose a form such as *this* or *that*; and so on. The GH delineates six hierarchically nested tiers of cognitive status (In Focus, Activated, Familiar, Uniquely Identifiable, Referential, and Type Identifiable), each associated with a different set of referring forms.

We choose the GH from the pragmatic family of models discussed above because (1) the GH provides perhaps the closest account of the connection between referring forms and information status as mediated by cognitive structures and processes, (2) the GH provides a commonsense explanation for how a wide variety of referring forms are related to different information statuses, and (3) there has been recent promising work within the robotics community for developing computational cognitive models of *understanding* wide varieties of referring forms using the GH (Williams, Acharya, Schreitter, & Scheutz, 2016; Williams & Scheutz, 2019). Our second goal is to furthermore provide similar computational cognitive GH-theoretic language *generation* that could be similarly implemented into robotics and other situated autonomous agents. As such, it will now be helpful to discuss related work that has been formed on computationally modeling referential choice.

Computational Models of Referential Choice

Much of this work on computational modeling of the process of referential choice (Poesio, Stevenson, Eugenio, & Hitzeman, 2004; McCoy & Strube, 1999; Ge, Hale, & Charniak, 1998; Kibrik et al., 2016; Kibrik, 2011; Callaway & Lester, 2002; Kibble & Power, 2004) falls under the broad categorization of multi-factorial process modeling, with selection of referential form viewed as a problem of classification based on animacy, grammatical role, and factors related to discourse structure, coherence, and salience. A discussion of the various predictive features that have been proposed and the machine learning approaches that have made use of these features, are surveyed by Kibrik (2011) (see also Van Deemter et al. (2012) and Gatt et al. (2014)).

These models have primarily made classification decisions between a small number of classes, such as pronoun vs description, or pronoun vs proper noun vs description, on the basis of large numbers of annotated features. This approach is directly related to the non-situated textual domains in which previous approaches have focused. First, previous approaches have been able to focus on simple classification decisions such as pronouns vs proper nouns vs descriptions as textual domains such as news reports in which there are a relatively small number of candidate referents, many of which are readily discriminable and which can be uniquely picked out using proper nouns. Second, previous approaches have made predictive decisions based on the large number of annotated features that are readily available in large-scale textual corpora.

In comparison, in *situated* domains present a number of complications. First, there are typically a large number of highly similar candidate referents, few of which can be referred to using proper names. For example, many previous applications have sought to model referential choice using corpora of Wall Street Journal articles, in which a given article may revolve around a small number of people, governments, institutions, and so forth, each of which can be readily referred to through a proper name, and about which a host of information has been annotated (Krasavina & Chiarcos, 2007). In contrast, in situated domains, speakers may need to distinguish between a large number of highly similar task-relevant entities (e.g., when collaboratively loading a dishwasher or setting a table, interlocutors may need to distinguish between many cups, plates, and so forth, which may be functionally identical and which are unlikely to have proper names). Moreover, robots operating in these domains must make classification decisions on the basis of features they can automatically assess, rather than the wider set of features that linguists can annotate. Moreover, in situated domains, the need for careful selection between ambiguous referring forms becomes arguably more important, both because certain referring forms (e.g., "this" and "that") are differentially used in situated contexts based on inherently situated features such as physical distance (as well as nonverbal cues such as gaze and gesture, although we do not consider them in this work either), and because overly ambiguous utterances in situated domains leads to expensive repair dialogues.

It is also valuable to consider the features used in previous computational cognitive models of referential choice, and their relation to linguistic theories of reference. While theories such as the Givenness Hierarchy have been widely successful in explaining from a linguistic perspective how the use of different referring forms can be motivated by speakers' presumptions regarding their referents' cognitive status, there has been little work seeking to explain the cognitive mechanisms and psycholinguistic processes by which cognitive status is used during language generation (Arnold, 2016), and neither has cognitive status been used as a feature in prior models of referential choice. This is natural, as cognitive status is a concept that exists in the mind of speakers, and is thus difficult to definitively annotate.

However, recent work in the cognitive science literature has begun to explore methods by which presumed cognitive status can be assessed from user judgments in human subject experiments, and how data from such experiments can thus be used to train predictive models to automatically predict presumed cognitive status (Pal et al., 2020). In this work, we leverage this recent work in order to directly predict referential choice from presumed cognitive status, along with a small number of theoretically-informed features of critical importance for situated interaction. Due to linguistic evidence that pragmatic models of referential choice are most naturally compatible with rule-based cognitive mechanisms, and due to the long history of linguists seeking to model referential choice using rule-based decision procedures (Levelt, 1993), the data-driven model we present in this work is learned using a highly explainable rule-based decision tree framework, to produce a model that is at once easily implementable in autonomous systems yet also readily mineable for linguistic and psycholinguistic insights.

Materials

Before describing our computational modeling approach, we will briefly describe the data modeled by our approach.

Situated Interaction Corpus

We used the dataset collected by Bennett et al. (2017), which involves a human instructor collaborating with human and robot learners in a spatially situated collaborative task, in the environment shown in Fig. 1. This environment contained four boxes labeled A-D (imposing some ambiguity despite the name-enabling label), three colored towers (yellow, red, blue), four walls, and a set of tapelines dividing the room into four quadrants. For simplicity, only the boxes and towers were annotated in this dataset.

The dataset consists of transcripts of videos of human participants (instructors) instructing human and robot teammates (learners) to rearrange the experimental environment, step by step, to achieve a reconfiguration of their choosing. Each participant performed this task twice, once with a human and once with a robot, with order counterbalanced. In both cases, the learner followed directions without responding. The dataset contains 66 monologues from 33 participants, each containing between 5 to 24 utterances referring to boxes or towers, for a total of 485 such utterances and a total of 603 such referring expressions. Two sample monologues are included below:

Monologue 1
So in front of you will be a box.
Hm, can you grab the second box?
You can move the box, actually, don't move you.
Take your hands, grab out, outside the box.
Just push it.
Now push that box over till I say stop.
Align with the box a little bit more.

Monologue 2
Um, first, we're gonna go push block A to the back-center of box 1.
Um, and knock over the blue tower.
Then we're gonna take box B and move it to the center of the 3rd quadrant.
Um, box C is going to go right on the number 4.
Um, and then box D is going to go to the corner on the line with box 1 on it, up against the wall.
Um, box C needs to be on the number 4.
Um, box D needs to be closer to the line.

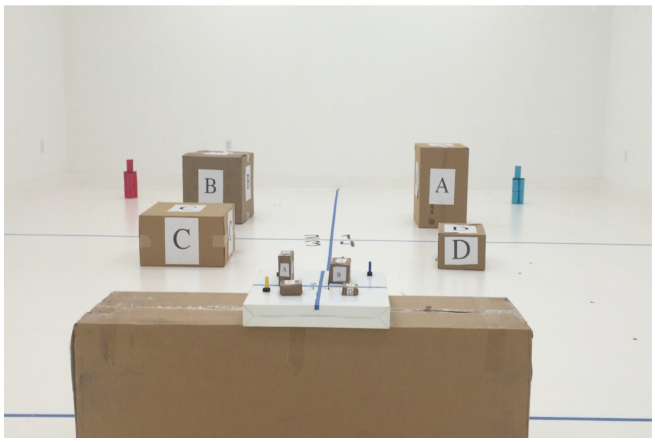

Figure 1: Situated Interaction Context (Bennett et al., 2017)

Classes

Each referring expression in this dataset was assigned a class label corresponding to the referring form used in that expression. Six class labels were used: *it, this, that, this* $\langle N' \rangle$*, that* $\langle N' \rangle$*,* and *the* $\langle N' \rangle$. The few indefinite noun phrases occurring in the dataset were not coded, leaving handling of such expressions to future work. Because this work sought to handle the main categories of referring forms handled by the GH, and these main categories do not include bare NPs, we take for convenience sake a descriptivist view (Frege, 1892; Russell, 2001; Nelson, 2002) and code bare noun phrases such as "Box A" as definite noun phrases (i.e., *the* $\langle N' \rangle$), leaving the distinction between definite descriptions and pseudo-proper names to future work.

Features

Each referring expression was annotated using cognitive status as well as three additional features well acknowledged to be critical to situated description: number of distractors, physical distance, and temporal distance. As discussed above, previous approaches to computational modeling of multi-factorial referential choice have tended to use large numbers of features that are readily available in large-scale annotated text corpora. In this work, however, we instead elected to focus on a small number of features of high theoretical relevance to the speaker's situated context that could be extracted on-line by situated autonomous agents such as robots. This approach should provide a model that is readily interpretable, efficient, and unlikely to be the result of overfitting to feature noise. We discuss each feature below.

Cognitive Status: The central thesis of this work is that referential choice may be more effectively modeled by primarily relying on the *cognitive status* referents are expected to hold in the mind of their interlocutors or in their current conversation (depending on one's theoretical interpretation of the Givenness Hierarchy). In order to use this as a feature in our computational models, each monologue was provided as input to a Cognitive Status Engine (Pal et al., 2020), comprised of a set of *Cognitive Status Filters*: Bayes Filters of the form:

$$p(S_o^t) = p(S_o^{t-1})p(L_o^t)p(S_o^t \mid S_o^{t-1}, L_o^t)$$

Each such cognitive status filter recursively estimates, for a given object o, the probability distribution over cognitive statuses S for o at time t, on the basis of linguistic features L. To estimate these distributions for a set of known objects, $O = \{o_1, ..., o_n\}$ at each time step, we use a Cognitive Status Modeling Engine C, consisting of a set of CSFs $\{c_0, ..., c_1\}$, one for each object believed to be of a status familiar or higher within the conversation.

In this work we used a CSE with one CSF for each task-relevant object. Using this CSE, we annotated each referring expression with the most likely cognitive status for its target referent (per its associated CSF) at the time of its utterance.

Number of Distractors: Much evidence suggests that speakers respond to the presence of distractors when speaking, avoiding pronouns when they would be ambiguous (Ferreira, Slevc, & Rogers, 2005), providing evidence for pragmatic models of referential choice (Ariel, 2014; Chafe, 1976; Givón, 1983; Gundel et al., 1993). However, while the presence of distractors is considered to be a determinant of accessibility (Ariel, 2014) and topicality (Givón, 1983), it is not viewed as a determinant of givenness, but rather as a factor that interacts with givenness when determining referential choice. As such, we include the number of distractors as a key feature in our model of referential choice. The number of distractors for a target referent was calculated as the number of other towers and boxes whose most likely cognitive status was estimated to be of the same GH-theoretic tier or higher as the target referent at the time of utterance, using the cognitive status estimation procedure discussed above.

Physical Distance: Due to our focus on situated contexts, we also included physical distance from speaker to target referent. There is substantial prior evidence of the role of physical distance in selecting between demonstratives such as "this" and "that" (Dixon, 2003). We encoded distance through video analysis of Bennett et al. (2017)'s corpus. Referents were classified as "close", "mid-distance", or "far" based on whether they were before, at, or beyond the most salient horizontal landmark (a line running the width of the room at roughly half the maximum distance into the room). While this means of classification is obviously tailored to this specific evaluation corpus, the general approach of considering whether a referent is in the nearer or farther half of a given task context is one that may easily generalize.

Temporal Distance: Finally, like most previous works on computational cognitive modeling of multi-factorial referential choice (Kibrik et al., 2016), we include a measure of linear or temporal distance, i.e., recency of mention (Givón, 1983). Specifically, we encode recency of mention for a target referent as $1/n$, where n is the number of referring expressions since the last mention of the target, with 1 meaning the target was the most recently mentioned object, and 0 meaning it has not yet been mentioned in the dialogue.

Computational Modeling

In this section we describe our computational cognitive model. We approach GH-informed referential choice as a classification problem of predicting class labels from input features. To solve this problem we chose to use the REPTree (Reduced Error Pruning Tree) implementation (Quinlan, 1987) from the open source WEKA software package (Hall et al., 2009). REPTrees are an extension of the classic C4.5 decision tree algorithm that builds a decision tree using an information gain based splitting criteria which is then pruned using a reduced-error pruning technique (Witten & Frank, 2002; Quinlan, 1993). Like previous approaches to computational cognitive modeling of multi-factorial referential choice, we selected a decision tree approach due to the ready interpetability of such approaches, which enables them to achieve reasonable accuracy while facilitating theory-building.

Five REPTree models were trained: a complete model (M1), and four ablated models removing either cognitive status (M2) distractors (M3), physical distance (M4), or temporal distance (M5). Models were evaluated via accuracy, root mean squared error (RMSE), precision, recall, and F1 score. All five models were trained by WEKA using stratified 15-fold cross-validation. Stratified evaluation was used to account for severe class imbalance between referring forms. The increased number of folds was chosen based on the preferences of the lead author.

Results

As shown in Tab. 1, all models achieved high accuracy, with the M5 model (excluding only temporal distance) achieving the best scores across primary metrics. The overall accuracy range for the top-performing models of 83-86% for six-class classification on the basis of four features is highly competitive with other recent models, which have received accuracy in the range of 72-75% for three-class classification on the basis of a greater number of features (Kibrik et al., 2016).

We compare to Kibrik et al. here because their work is most similar to our own. Critically, however, their work examined a substantially different (non-situated) domain. Moreover, they sought to model the choice between proper names, descriptions, and pronouns, which is an overlapping set of classes to our own but not a strict subset. As such, our comparison here is not intended as formal evidence of "greater" performance of our approach, but merely a signifier of comparable results to the most similar prior work.

Our top performing models performed roughly equivalently; selection between these models can be used on the basis of other factors, such as those shown in the last two rows of Tab. 1: coverage (modeled as number of classes included in model predictions) and model simplicity (modeled as number of leaves). Intuitively, a model should be able to predict the use of all referring forms included as class labels, without being overly complex (a sign of overfitting). Our results show that only the full model (despite its high number of leaves) accounted for all referring forms included as class labels. As such, we analyze only this full model.

Table 1: Evaluation Metrics

	M1	M2	M3	M4	M5
Accuracy	84.74	79.6	71.97	83.58	**86.07**
RMSE	0.197	0.230	0.244	0.208	**0.195**
Precision	0.858	0.820	0.710	0.840	**0.882**
Recall	0.847	0.796	0.720	0.836	**0.861**
F1 score	0.843	0.811	0.716	0.838	**0.858**
Coverage	6/6	5/6	3/6	3/6	3/6
Leaves	13	9	5	9	6

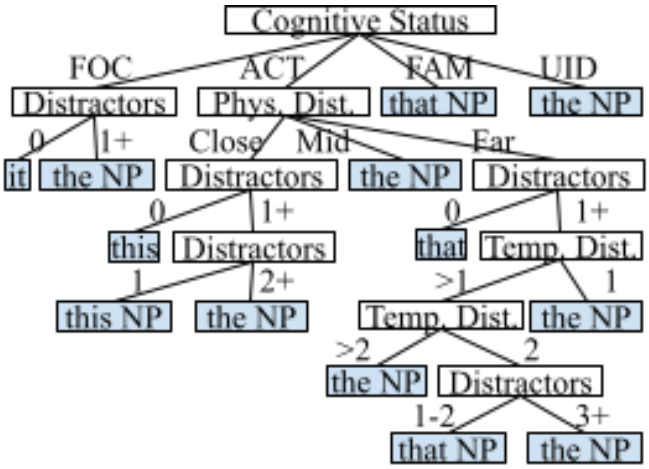

Figure 2: Unablated (M1) Model. FOC = In Focus; ACT = Activated; FAM = Familiar; UID = Uniquely Identifiable

As shown in Fig. 2, the M1 model can be interpreted as follows. First, the model begins by considering the cognitive status of the target referent. If the target is at most uniquely identifiable, the model selects *the* ⟨*N'*⟩. If the target is at most familiar, the model selects *that* ⟨*N'*⟩. If the target is at most activated, a substantially more complicated chain of reasoning is performed, as described below. Finally, if the target is at most in focus, then if there are no other referents in focus, the model selects *it*, and otherwise selects *the* ⟨*N'*⟩. The jump directly from pronoun to definite description is interesting, as the model does not even consider forms like *this* or *that*, which would also be felicitous. This is perhaps because those forms would within a pragmatic account signal the target to be at most activated rather than in focus.

When the target is at most activated, the model first considers physical distance to the referent. If the target is neither close nor far, the model chooses *the* ⟨*N'*⟩. If close, the decision depends on number of distractors: If there are no other activated entities, the model chooses *this*; if there is a single activated distractor, the model chooses *this* ⟨*N'*⟩; otherwise the model chooses *the* ⟨*N'*⟩. If far, the decision depends on both number of distractors and temporal distance: if there is

a single familiar distractor, the model chooses *that*; otherwise if there is a single other entity that was mentioned more recently, and there are only 1-2 distractors, the model chooses *that* ⟨*N'*⟩; otherwise the model chooses *the* ⟨*N'*⟩.

Discussion and Conclusion

Our results demonstrate the efficacy of cognitive status as a practical means of predicting referential choice in situated contexts, especially when combined with other features critical to situated contexts, such as speaker-referent distance and number of distractors of equivalent status. For the psycholinguistics (and linguistics) communities, these results support previous arguments framing referential choice as a rule-based process guided by mental assumptions regarding cognitive status, and present a straightforward and easily interpretable model of the *cognitive process* of referential choice. For the artificial intelligence community, our results represent a straightforward model that can easily be integrated into robot cognitive architectures to guide referential choice, which will allow for more accurate and natural human-robot interaction. This work also raises a number of questions and directions which we are seeking to address in ongoing and future work.

First, while our model predicts a wider range of referring forms than previous work, it does not cover the complete diversity of referring forms, including bare noun phrases, indefinite noun phrases, personal pronouns, and so forth. The success of the approach presented in this work suggests that this approach would likely also be successful when trained on an expanded dataset with a larger number of annotated classes. Doing so, however, will require collection of a larger dataset in which this broader variety of referring forms are used.

Second and similarly, our training corpus was limited to a few objects, and did not include references to environmental geometry, events, people, and so forth. Handling these additional entities would likely further increase our model's performance, due to a more accurate modeling of the set of distractors.

Third, while the model predicts referring forms on the basis of cognitive status, it does not account for the mechanisms by which humans use gaze and gesture to manipulate the status of entities. E.g., a referent may not be activated at the time a speaker decides to refer to it, but it may be activated by the time the user commences speaking. This suggests that the model presented in this work may benefit from being embedded within a larger strategic model of this form.

Fourth, while the key predictive feature of our model was cognitive status, the model of cognitive status we used was not itself perfectly accurate. As discussed by Pal et al. (2020), their Cognitive Status Filtering accuracy is around 82%; as such, a refined model of cognitive status that itself uses a greater number of predictive features to enhance accuracy would likely result in equivalently increased accuracy for our own model of referential choice.

Fifth, while the choice of decision tree model was not of significant interest to us in this work, in future work we plan to compare the REPTrees used in this work to more commonplace or modern alternatives such as J48 and XGBoost (Chen & Guestrin, 2016).

Finally, similar to what has been noted by Kibrik et al. (2016), while models of multi-factorial referential choice may predict individual referential choices, in a given situation many choices may be equally acceptable, and choice between those options may be a matter of speaker preference, especially given the range of referring forms considered in this work. This means that "incorrect" predictions made by our model may not actually be problematic and in need of improvement. This suggests the need for task-based evaluations of the effectiveness and naturality of selected forms, similar to evaluations performed by Kibrik et al. (2016), Williams and Scheutz (2017), and van der Lee et al. (2020).

Acknowledgments

This work relates to Department of Navy award N00014-21-1-2418 issued by the Office of Naval Research.

References

Ariel, M. (1991). The function of accessibility in a theory of grammar. *Journal of pragmatics*, *16*(5), 443–463.

Ariel, M. (2014). *Accessing noun-phrase antecedents*.

Arnold, J. E. (2016). Explicit and emergent mechanisms of information status. *Topics in Cog. Sci.*.

Arnold, J. E., & Zerkle, S. A. (2019). Why do people produce pronouns? pragmatic selection vs. rational models. *Language, Cognition and Neuroscience*, *34*(9), 1152–1175.

Aylett, M., & Turk, A. (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and speech*, *47*(1), 31–56.

Bennett, M., Williams, T., Thames, D., & Scheutz, M. (2017). Differences in interaction patterns and perception for teleoperated and autonomous humanoid robots. In *IROS*.

Brennan, S. E., Friedman, M. W., & Pollard, C. (1987). A centering approach to pronouns. In *ACL*.

Callaway, C. B., & Lester, J. C. (2002). Pronominalization in generated discourse and dialogue. In *ACL*.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics, and point of view. *Subject and topic*.

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794).

Dixon, R. M. (2003). Demonstratives: A cross-linguistic typology. *Studies in Language*.

Ferreira, V., Slevc, L., & Rogers, E. (2005). How do speakers avoid ambiguous linguistic expressions? *Cognition*, *96*(3).

Fox, B. A. (1993). *Discourse structure and anaphora: Written and conversational english* (No. 48).

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*(6084).

Frege, G. (1892). Über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*.

Gatt, A., & Krahmer, E. (2018). Survey of the state of the art in natural language generation: Core tasks, applications and evaluation. *Journal of Artificial Intelligence Research*.

Gatt, A., Krahmer, E., Van Deemter, K., & Van Gompel, R. (2014). Models and empirical data for the production of referring expressions. *Lang., Cognition and Neuroscience*.

Ge, N., Hale, J., & Charniak, E. (1998). A statistical approach to anaphora resolution. In *Workshop on very large corpora.*

Givón, T. (1983). *Topic continuity in discourse: A quantitative cross-language study* (Vol. 3).

Grosz, B. J., Joshi, A. K., & Weinstein, S. (1995). Centering: A framework for modelling the local coherence of discourse. *Computational Linguistics*.

Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*.

Grüning, A., & Kibrik, A. A. (2005). Modelling referential choice in discourse: A cognitive calculative approach and a neural network approach. *Anaphora processing: Linguistic, cognitive and computational modelling*, *263*, 163.

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

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*.

Hedberg, N. (2013). Applying the givenness hierarchy framework: Methodological issues. *International workshop on information structure of Austronesian languages*.

Kibble, R., & Power, R. (2004). Optimizing referential coherence in text generation. *Comp. Ling.*.

Kibrik, A. A. (2011). *Reference in discourse*. OUP.

Kibrik, A. A., Khudyakova, M. V., Dobrov, G. B., Linnik, A., & Zalmanov, D. A. (2016). Referential choice: Predictability and its limits. *Frontiers in psychology*, *7*, 1429.

Krahmer, E., & Van Deemter, K. (2012). Computational generation of referring expressions: A survey. *Comp. Ling.*.

Krasavina, O., & Chiarcos, C. (2007). Pocos-potsdam coreference scheme. In *Linguistic annotation workshop.*

Levelt, W. (1993). *Speaking: From intention to articulation*.

Mann, W. C., Matthiessen, C. M., & Thompson, S. A. (1989). *Rhetorical structure theory and text analysis* (Tech. Rep.). University of Southern California.

Marslen-Wilson, W., Levy, E., & Tyler, L. K. (1982). Producing interpretable discourse: The establishment and maintenance of reference. *Speech, place, and action*, 339–378.

McCoy, K. F., & Strube, M. (1999). Generating anaphoric expressions: pronoun or definite description? In *The relation of discourse/dialogue structure and reference.*

Nelson, M. (2002). Descriptivism defended. *Noûs*.

Pal, P., Zhu, L., Golden-Lasher, A., Swaminathan, A., & Williams, T. (2020). Givenness hierarchy theoretic cognitive status filtering. In *Cogsci* (pp. 925–931).

Poesio, M., Stevenson, R., Eugenio, B. D., & Hitzeman, J. (2004). Centering: A parametric theory and its instantiations. *Computational linguistics*, *30*(3), 309–363.

Quinlan, J. (1987). Simplifying decision trees. *International journal of man-machine studies*, *27*(3), 221–234.

Quinlan, J. (1993). Program for machine learning. *C4. 5*.

Rosa, E. C., & Arnold, J. E. (2011). The role of attention in choice of referring expressions. *PRE-Cogsci: Bridging computational, emp. and theoretical app. to reference*.

Russell, B. (2001). *The problems of philosophy*. OUP.

Tomlin, R. S. (1987). Linguistic reflections of cognitive events. *Coherence and grounding in discourse*, *11*.

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

Van Deemter, K., Gatt, A., Van Gompel, R. P., & Krahmer, E. (2012). Toward a computational psycholinguistics of reference production. *Topics in cognitive science*.

van der Lee, C., Gatt, A., van Miltenburg, E., & Krahmer, E. (2020). Human evaluation of automatically generated text: Current trends and best practice guidelines. *Computer Speech & Language*.

Williams, T., Acharya, S., Schreitter, S., & Scheutz, M. (2016). Situated open world reference resolution for human-robot dialogue. In *HRI.*

Williams, T., & Scheutz, M. (2017). Referring expression generation under uncertainty: Algorithm and evaluation framework. In *INLG.*

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

Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with java implementations. *ACM SIGMOD Record*, *31*(1), 76–77.