To appear in: Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue. Queen Mary University of London, Sep 11-12.

Not a simple yes or no: Uncertainty in indirect answers

Marie-Catherine de Marneffe, Scott Grimm and **Christopher Potts**
Linguistics Department
Stanford University
Stanford, CA 94305
{mcdm,sgrimm,cgpotts}@stanford.edu

Abstract

There is a long history of using logic to model the interpretation of indirect speech acts. Classical logical inference, however, is unable to deal with the combinations of disparate, conflicting, uncertain evidence that shape such speech acts in discourse. We propose to address this by combining logical inference with probabilistic methods. We focus on responses to polar questions with the following property: they are neither *yes* nor *no*, but they convey information that can be used to infer such an answer with some degree of confidence, though often not with enough confidence to count as resolving. We present a novel corpus study and associated typology that aims to situate these responses in the broader class of indirect question–answer pairs (IQAPs). We then model the different types of IQAPs using Markov logic networks, which combine first-order logic with probabilities, emphasizing the ways in which this approach allows us to model inferential uncertainty about both the context of utterance and intended meanings.

1 Introduction

Clark (1979), Perrault and Allen (1980), and Allen and Perrault (1980) study indirect speech acts, identifying a wide range of factors that govern how speakers convey their intended messages and how hearers seek to uncover those messages. Prior discourse conditions, the relationship between the literal meaning and the common ground, and specific lexical, constructional, and intonational cues all play a role. Green and Carberry (1992, 1994) provide an extensive computational model that interprets and generates indirect answers to polar questions. Their model focuses on inferring categorical answers, making use of discourse plans and coherence relations.

This paper extends such work by recasting the problem in terms of probabilistic modeling. We focus on the interpretation of indirect answers where the respondent does not answer with *yes* or *no*, but rather gives information that can be used by the hearer to infer such an answer only with some degree of certainty, as in (1).

(1) A: Is Sue at work?
B: She is sick with the flu.

In this case, whether one can move from the response to a *yes* or *no* is uncertain. Based on typical assumptions about work and illness, A might take B's response as indicating that Sue is at home, but B's response could be taken differently depending on Sue's character — B could be reproaching Sue for her workaholic tendencies, which risk infecting the office, or B could be admiring Sue's steadfast character. What A actually concludes about B's indirect reply will be based on some combination of this disparate, partially conflicting, uncertain evidence. The plan and logical inference model of Green and Carberry falters in the face of such collections of uncertain evidence. However, natural dialogues are often interpreted in the midst of uncertain and conflicting signals. We therefore propose to enrich a logical inference model with probabilistic methods to deal with such cases.

This study addresses the phenomenon of indirect question–answer pairs (IQAP), such as in (1), from both empirical and engineering perspectives.

First, we undertake a corpus study of polar questions in dialogue to gather naturally occurring instances and to determine how pervasive indirect answers that indicate uncertainty are in a natural setting (section 2). From this empirical base, we provide a classification of IQAPs which makes a new distinction between fully- and partially-resolving answers (section 3). We then show how inference in Markov logic networks can successfully model the reasoning involved in both types of IQAPs (section 4).

2 Corpus study

Previous corpus studies looked at how pervasive indirect answers to yes/no questions are in dialogue. Stenström (1984) analyzed 25 face-to-face and telephone conversations and found that 13% of answers to polar questions do not contain an explicit *yes* or *no* term. In a task dialogue, Hockey et al. (1997) found 38% of the responses were IQAPs. (This higher percentage might reflect the genre difference in the corpora used: task dialogue vs. casual conversations.) These studies, however, were not concerned with how confidently one could infer a *yes* or *no* from the response given.

We therefore conducted a corpus study to analyze the types of indirect answers. We used the Switchboard Dialog Act Corpus (Jurafsky et al., 1997) which has been annotated for approximately 60 basic dialog acts, clustered into 42 tags. We are concerned only with direct yes/no questions, and not with indirect ones such as "May I remind you to take out the garbage?" (Clark, 1979; Perrault and Allen, 1980). From 200 5-minute conversations, we extracted yes/no questions (tagged "qy") and their answers, but discarded tag questions as well as disjunctive questions, such as in (2), since these do not necessarily call for a *yes* or *no* response. We also did not take into account questions that were lost in the dialogue, nor questions that did not really require an answer (3). This yielded a total of 623 yes/no questions.

(2) [sw_0018_4082]

A: Do you, by mistakes, do you mean just like honest mistakes

A: or do you think they are deliberate sorts of things?

B: Uh, I think both.

(3) [sw_0070_3435]

A: How do you feel about your game?

A: I guess that's a good question?

B: Uh, well, I mean I'm not a serious golfer at all.

To identify indirect answers, we looked at the answer tags. The distribution of answers is given in Table 1. We collapsed the tags into 6 categories. Category I contains direct yes/no answers as well as "agree" answers (e.g., *That's exactly it.*). Category II includes statement–opinion and statement–non-opinion: e.g., *I think it's great*, *Me I'm in the legal department*, respectively. Affirmative non-*yes* answers and negative non-*no* answers form category III. Other answers such as *I don't know* are in category IV. In category V, we put utterances that avoid answering the question: by holding (*I'm drawing a blank*), by returning the question — wh-question or rhetorical question (*Who would steal a newspaper?*) — or by using a backchannel in question form (*Is that right?*). Finally, category VI contains dispreferred answers (Schegloff et al., 1977; Pomerantz, 1984).

We hypothesized that the phenomenon we are studying would appear in categories II, III and VI. However, some of the "na/ng" answers are disguised yes/no answers, such as *Right*, *I think so*, or *Not really*, and as such do not interest us. In the case of "sv/sd" and "nd" answers, many answers include reformulation, question avoidance (see 4), or a change of framing (5). All these cases are not really at issue for the question we are addressing.

(4) [sw_0177_2759]

A: Have you ever been drug tested?

B: Um, that's a good question.

(5) [sw_0046_4316]

A: Is he the guy wants to, like, deregulate heroin, or something?

B: Well, what he wants to do is take all the money that, uh, he gets for drug enforcement and use it for, uh, drug education.

A: Uh-huh.

B: And basically, just, just attack the problem at the demand side.

	Definition	Tag	Total
I	yes/no answers	ny/nn/aa	341
II	statements	sv/sd	143
III	affirmative/negative non-yes/no answers	na/ng	91
IV	other answers	no	21
V	avoid answering	^h/qw/qh/bh	18
VI	dispreferred answers	nd	9
Total			623

Table 1: Distribution of answer tags to yes/no questions.

(6) [sw_0046_4316]

A: That was also civil?

B: The other case was just traffic, and you know, it was seat belt law.

We examined by hand all yes/no questions for IQAPs and found 88 examples (such as (6), and (7)–(11)), which constitutes thus 14% of the total answers to direct yes/no questions, a figure similar to those of Stenström (1984). The next section introduces our classification of answers.

3 Typology of indirect answers

We can adduce the general space of IQAPs from the data assembled in section 2 (see also Bolinger, 1978; Clark, 1979). One point of departure is that, in cooperative dialogues, a response to a question counts as an answer only when some relation holds between the content of the response and the semantic desiderata of the question. This is succinctly formulated in the relation *IQAP* proposed by Asher and Lascarides (2003), p. 403:

> IQAP(α,β) holds only if there is a true direct answer p to the question $[\![\alpha]\!]$, and the questioner can infer p from $[\![\beta]\!]$ in the utterance context.

The apparent emphasis on truth can be set aside for present purposes; Asher and Lascarides's notions of truth are heavily relativized to the current discourse conditions. This principle hints at two dimensions of IQAPs which must be considered, and upon which we can establish a classification: (i) the type of answer which the proffered response provides, and (ii) the basis on which the inferences are performed. The typology established here adheres to this, distinguishing between fully- and partially-resolving answers as well as between the types of knowledge used in the inference (logical, linguistic, common ground/world).

3.1 Fully-resolving responses

An indirect answer can fully resolve a question by conveying information that stands in an inclusion relation to the direct answer: if $q \subseteq p$ (or $\neg p$), then updating with the response q also resolves the question with p (or $\neg p$), assuming the questioner knows that the inclusion relation holds between q and p. The inclusion relation can be based on logical relations, as in (7), where the response is an "over-answer", i.e., a response where more information is given than is strictly necessary to resolve the question. Hearers supply more information than strictly asked for when they recognize that the speaker's intentions are more general than the question posed might suggest. In (7), the most plausible intention behind the query is to know more about B's family. The hearer can also identify the speaker's plan and any necessary information for its completion, which he then provides (Allen and Perrault, 1980).

(7) [sw_0001_4325]

A: Do you have kids?

B: I have three.

While logical relations between the content of the question and the response suffice to treat examples such as (7), other over-answers often require substantial amounts of linguistic and/or world-knowledge to allow the inference to go through, as in (8) and (9).

(8) [sw_0069_3144]

A: Was that good?

B: Hysterical. We laughed so hard.

(9) [sw_0057_3506]

A: Is it in Dallas?

B: Uh, it's in Lewisville.

In the case of (8), a system must recognize that *hysterical* is semantically stronger than *good*. Similarly, to recognize the implicit *no* of (9), a system must recognize that Lewisville is a distinct location from Dallas, rather than, say, contained in Dallas, and it must include more general constraints as well (e.g., an entity cannot be in two physical locations at once). Once the necessary knowledge is in place, however, the inferences are properly licensed.

3.2 Partially-resolving responses

A second class of IQAPs, where the content of the answer itself does not fully resolve the question, known as partially-resolved questions (Groenendijk and Stokhof, 1984; Zeevat, 1994; Roberts, 1996; van Rooy, 2003), is less straightforward. One instance is shown in (10), where the gradable adjective *little* is the source of difficulty.

(10) [sw_0160_3467]

A: Are they [your kids] little?

B: I have a seven-year-old and a ten-year-old.

A: Yeah, they're pretty young.

The response, while an answer, does not, in and of itself, resolve whether the children should be considered *little*. The predicate *little* is a gradable adjective, which inherently possesses a degree of vagueness: such adjectives contextually vary in truth conditions and admit borderline cases (Kennedy, 2007). In the case of *little*, while some children are clearly little, e.g., ages 2–3, and some clearly are not, e.g., ages 14–15, there is another class in between for which it is difficult to assess whether *little* can be truthfully ascribed to them. Due to the slippery nature of these predicates, there is no hard-and-fast way to resolve such questions in all cases. In (10), it is the questioner who resolves the question by accepting the information proffered in the response as sufficient to count as *little*.

The dialogue in (11) shows a second example of an answer which is not fully-resolving, and intentionally so.

(11) [sw_0103_4074]

A: Did he raise him [the cat] or something[1]?

B: We bought the cat for him and so he's been the one that you know spent the most time with him.

Speaker B quibbles with whether the relation his son has to the cat is one of *raising*, instead citing two attributes that go along with, but do not determine, *raising*. *Raising an animal* is a composite relation, which typically includes the relations *owning* and *spending time with*. However, satisfying these two sub-relations does not strictly entail satisfying the *raising* relation as well. It is not obvious whether a system would be mistaken in attributing a fully positive response to the question, although it is certainly a *partially* positive response. Similarly, it seems that attributing a negative response would be misguided, though the answer is partly negative. The rest of the dialogue does not determine whether A considers this equivalent to *raising*, and the dialogue proceeds happily without this resolution.

The preceding examples have primarily hinged upon conventionalized linguistic knowledge, viz. what it means to *raise X* or for *X* to be *little*. A further class of partially-resolving answers relies on knowledge present in the common ground. Our initial example (1) illustrates a situation where different resolutions of the question were possible depending on the respondent's intentions: *no* if sympathetic, *yes* if reproachful or admiring.

The relationship between the response and question is not secured by any objective world facts or conventionalized meaning, but rather is variable — contingent on specialized world knowledge concerning the dialogue participants and their beliefs. Resolving such IQAPs positively or negatively is achieved only at the cost of a degree of uncertainty: for resolution occurs against the backdrop of a set of defeasible assumptions.

3.3 IQAP classification

Table 2 is a cross-classification of the examples discussed by whether the responses are fully- or partially-resolving answers and by the types of knowledge used in the inference (logical, linguistic, world). It gives, for each category, the counts of examples we found in the corpus. The partially-resolved class contains more than a third of the answers.

[1]The disjunct *or something* may indicate that A is open to hearing about alternatives to *raise*. We abstract away from this issue for present purposes and treat the more general case by assuming A's contribution is simply equivalent to "Did he raise him?"

	Logic	Linguistic	World	Total
Fully-Resolved	27 (Ex. 7)	18 (Ex. 8)	11 (Ex. 9)	56
Partially-Resolved	–	20 (Ex. 10;11)	12 (Ex. 1)	32

Table 2: Classification of IQAPs by knowledge type and resolvedness: counts and examples.

The examples given in (7)–(9) are fully resolvable via inferences grounded in logical relations, linguistic convention or objective facts: the answer provides enough information to fully resolve the question, and the modeling challenge is securing and making available the correct information. The partially-resolved pairs are, however, qualitatively different. They involve a degree of uncertainty that classical inference models do not accommodate in a natural way.

4 Towards modeling IQAP resolution

To model the reasoning involved in all types of IQAPs, we can use a relational representation, but we need to be able to deal with uncertainty, as highlighted in section 3. Markov logic networks (MLNs; Richardson and Domingos, 2006) exactly suit these needs: they allow rich inferential reasoning on relations by combining the power of first-order logic and probabilities to cope with uncertainty. A logical knowledge-base is a set of hard constraints on the set of possible worlds (set of constants and grounded predicates). In Markov logic, the constraints are "soft": when a world violates a relation, it becomes less probable, but not impossible. A Markov logic network encodes a set of weighted first-order logic constraints, such that a higher weight implies a stronger constraint. Given constants in the world, the MLN creates a network of grounded predicates which applies the constraints to these constants. The network contains one feature f_j for each possible grounding of each constraint, with a value of 1 if the grounded constraint is true, and 0 otherwise. The probability of a world x is thus defined in terms of the constraints j satisfied by that world and the weights w associated with each constraint (Z being the partition function):

$$P(X = x) = \frac{1}{Z} \sum_j w_j f_j(x)$$

In practice, we use the Alchemy implementation of Markov logic networks (Kok et al., 2009). Weights on the relations can be hand-set or learned. Currently, we use weights set by hand, which suffices to demonstrate that an MLN handles the pragmatic reasoning we want to model, but ultimately we would like to learn the weights.

In this section, we show by means of a few examples how MLNs give a simple and elegant way of modeling the reasoning involved in both partially- and fully-resolved IQAPs.

4.1 Fully-resolved IQAPs

While the use of MLNs is motivated by partially-resolved IQAPs, to develop the intuitions behind MLNs, we show how they model fully-resolved cases, such as in (9). We define two distinct places, Dallas and Lewisville, a relation linking a person to a place, and the fact that person K is in Lewisville. We also add the general constraint that an individual can be in only one place at a time, to which we assign a very high weight. Markov logic allows for infinite weights, which Alchemy denotes by a closing period. We also assume that there is another person L, whose location is unknown.

Constants and facts:
Place = {Dallas, Lewisville}
Person = {K,L}
BeIn(Person,Place)
BeIn(K,Lewisville)

Constraints:
// "If you are in one place, you are not in another."
(BeIn(x,y) ∧ (y != z)) ⇒ !BeIn(x,z).

Figure 1 represents the grounded Markov network obtained by applying the constraint to the constants K, L, Dallas and Lewisville. The graph contains a node for each predicate grounding, and an arc between each pair of nodes that appear together in some grounding of the constraint. Given that input, the MLN samples over possible worlds, and infers probabilities for the predicate BeIn, based on the constraints satisfied by each world and their weights. The MLN returns a very low probability for K being in Dallas, meaning that the answer to the question *Is it in Dallas?* is *no*:

BeIn(K,Dallas): 4.9995e-05

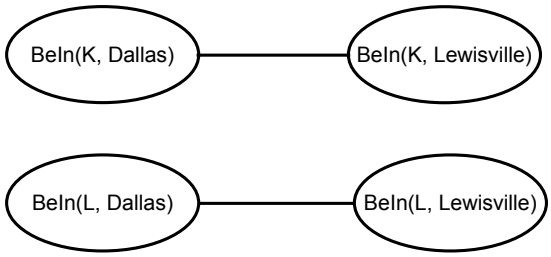

Figure 1: Grounded Markov network obtained by applying the constraints to the constants K, L, Dallas and Lewisville.

Since no information about L's location has been given, the probabilities of L being in Dallas or Lewisville will be equal and low (0.3), which is exactly what one would hope for. The probabilities returned for each location will depend on the number of locations specified in the input.

4.2 Partially-resolved IQAPs

To model partially-resolved IQAPs appropriately, we need probabilities, since such IQAPs feature reasoning patterns that involve uncertainty. We now show how we can handle three examples of partially-resolved IQAPs.

Gradable adjectives. Example (10) is a borderline case of gradable adjectives: the question bears on the predicate *be little* for two children of ages 7 and 10. We first define the constants and facts about the world, which take into account the relations under consideration, "BeLittle(X)" and "Age(X, i)", and specify which individuals we are talking about, K and L, as well as their ages.

Constants and facts:
age = $\{0 \ldots 120\}$
Person = $\{$K, L$\}$
Age(Person,age)
BeLittle(Person)
Age(K,7)
Age(L,10)

The relation between age and being little involves some uncertainty, which we can model using a logistic curve. We assume that a 12-year-old child lies in the vague region for determining "littleness" and therefore 12 will be used as the center of the logistic curve.

Constraints:
// "If you are under 12, you are little."
1.0 (Age(x,y) $\wedge$ y $<$ 12) $\Rightarrow$ BeLittle(x)
// "If you are above 12, you are not little."
1.0 (Age(x,y) $\wedge$ y $>$ 12) $\Rightarrow$!BeLittle(x)
// The constraint below links two instances of BeLittle.
(Age(x,u) $\wedge$ Age(y,v) $\wedge$ v$>$u $\wedge$ BeLittle(y)) $\Rightarrow$ BeLittle(x).

Asking the network about K being little and L being little, we obtain the following results, which lead us to conclude that K and L are indeed little with a reasonably high degree of confidence, and that the indirect answer to the question is heavily biased towards *yes*.

BeLittle(K): 0.92
BeLittle(L): 0.68

If we now change the facts, and say that K and L are respectively 12 and 16 years old (instead of 7 and 10), we see an appropriate change in the probabilities:

BeLittle(K): 0.58
BeLittle(L): 0.16

L, the 16-year-old, is certainly not to be considered "little" anymore, whereas the situation is less clear-cut for K, the 12-year-old (who lies in the vague region of "littleness" that we assumed).

Ideally, we would have information about the speaker's beliefs, which we could use to update the constraints' weights. Absent such information, we could use general knowledge from the Web to learn appropriate weights. In this specific case, we could find age ranges appearing with "little kids" in data, and fit the logistic curve to these.

This probabilistic model adapts well to cases where categorical beliefs fit uneasily: for borderline cases of vague predicates (whose interpretation varies by participant), there is no deterministic *yes* or *no* answer.

Composite relations. In example (11), we want to know whether the speaker's son raised the cat inasmuch as he owned and spent time with him. We noted that *raise* is a composite relation, which entails simpler relations, in this case *spend time with* and *own*, although satisfying any one of the simpler relations does not suffice to guarantee the truth of *raise* itself. We model the constants, facts, and constraints as follows:

Constants and Facts:
Person = {K}
Animal = {Cat}
Raise(Person,Animal)
SpendTime(Person,Animal)
Own(Person,Animal)
SpendTime(K,Cat)
Own(K,Cat)

Constraints:
// "If you spend time with an animal, you help raise it."
1.0 SpendTime(x,y) $\Rightarrow$ Raise(x,y)
// "If you own an animal, you help raise it."
1.0 Own(x,y) $\Rightarrow$ Raise(x,y)

The weights on the relations reflect how central we judge them to be in defining *raise*. For simplicity, here we let the weights be identical. Clearly, the greater number of relevant relations a pair of entities fulfills, the greater the probability that the composite relation holds of them. Considering two scenarios helps illustrate this. First, suppose, as in the example, that both relations hold. We will then have a good indication that by owning and spending time with the cat, the son helped raise him:

Raise(K,Cat): 0.88

Second, suppose that the example is different in that only one of the relations holds, for instance, that the son only spent time with the cat, but did not own it, and accordingly the facts in the network do not contain Own(K,Cat). The probability that the son raised the cat decreases:

Raise(K,Cat): 0.78

Again this can easily be adapted depending on the centrality of the simpler relations to the composite relation, as well as on the world-knowledge concerning the (un)certainty of the constraints.

Speaker beliefs and common ground knowledge. The constructed question–answer pair given in (1), concerning whether Sue is at work, demonstrated that how an indirect answer is modeled depends on different and uncertain evidence. The following constraints are intended to capture some background assumptions about how we regard working, being sick, and the connections between those properties:

// "If you are sick, you are not coming to work."
Sick(x) $\Rightarrow$!AtWork(x)
// "If you are hardworking, you are at work."
HardWorking(x) $\Rightarrow$ AtWork(x)
// "If you are malicious and sick, you come to work."
(Malicious(x) $\wedge$ Sick(x)) $\Rightarrow$ AtWork(x)
// "If you are at work and sick, you are malicious or thoughtless."
(AtWork(x) $\wedge$ Sick(x)) $\Rightarrow$ (Malicious(x) $\vee$ Thoughtless(x))

These constraints provide different answers about Sue being at work depending on how they are weighted, even while the facts remain the same in each instance. If the first constraint is heavily weighted, we get a high probability for Sue not being at work, whereas if we evenly weight all the constraints, Sue's quality of being a hard-worker dramatically raises the probability that she is at work. Thus, MLNs permit modeling inferences that hinge upon highly variable common ground and speaker beliefs.

Besides offering an accurate treatment of fully-resolved inferences, MLNs have the ability to deal with degrees of certitude. This power is required if one wants an adequate model of the reasoning involved in partially-resolved inferences. Indeed, for the successful modeling of such inferences, it is essential to have a mechanism for adding facts about the world that are accepted to various degrees, rather than categorically, as well as for updating these facts with speakers' beliefs if such information is available.

5 Conclusions

We have provided an empirical analysis and initial treatment of indirect answers to polar questions. The empirical analysis led to a categorization of IQAPs according to whether their answers are fully- or partially-resolving and according to the types of knowledge used in resolving

the question by inference (logical, linguistic, common ground/world). The partially-resolving indirect answers injected a degree of uncertainty into the resolution of the predicate at issue in the question. Such examples highlight the limits of traditional logical inference and call for probabilistic methods. We therefore modeled these exchanges with Markov logic networks, which combine the power of first-order logic and probabilities. As a result, we were able to provide a robust model of question–answer resolution in dialogue, one which can assimilate information which is not categorical, but rather known only to a degree of certitude.

Acknowledgements

We thank Christopher Davis, Dan Jurafsky, and Christopher D. Manning for their insightful comments on earlier drafts of this paper. We also thank Karen Shiells for her help with the data collection and Markov logic.

References

James F. Allen and C. Raymond Perrault. 1980. Analyzing intention in utterances. *Artificial Intelligence*, 15:143–178.

Nicholas Asher and Alex Lascarides. 2003. *Logics of Conversation*. Cambridge University Press, Cambridge.

Dwight Bolinger. 1978. Yes–no questions are not alternative questions. In Henry Hiz, editor, *Questions*, pages 87–105. D. Reidel Publishing Company, Dordrecht, Holland.

Herbert H. Clark. 1979. Responding to indirect speech acts. *Cognitive Psychology*, 11:430–477.

Nancy Green and Sandra Carberry. 1992. Conversational implicatures in indirect replies. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, pages 64–71, Newark, Delaware, USA, June. Association for Computational Linguistics.

Nancy Green and Sandra Carberry. 1994. A hybrid reasoning model for indirect answers. In *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, pages 58–65, Las Cruces, New Mexico, USA, June. Association for Computational Linguistics.

Jeroen Groenendijk and Martin Stokhof. 1984. *Studies in the Semantics of Questions and the Pragmatics of Answers*. Ph.D. thesis, University of Amsterdam.

Beth Ann Hockey, Deborah Rossen-Knill, Beverly Spejewski, Matthew Stone, and Stephen Isard. 1997. Can you predict answers to Y/N questions? Yes, No and Stuff. In *Proceedings of Eurospeech 1997*.

Daniel Jurafsky, Elizabeth Shriberg, and Debra Biasca. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual, draft 13. Technical Report 97-02, University of Colorado, Boulder Institute of Cognitive Science.

Christopher Kennedy. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, 30(1):1–45.

Stanley Kok, Marc Sumner, Matthew Richardson, Parag Singla, Hoifung Poon, Daniel Lowd, Jue Wang, and Pedro Domingos. 2009. The Alchemy system for statistical relational AI. Technical report, Department of Computer Science and Engineering, University of Washington, Seattle, WA.

C. Raymond Perrault and James F. Allen. 1980. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3-4):167–182.

Anita M. Pomerantz. 1984. Agreeing and disagreeing with assessment: Some features of preferred/dispreferred turn shapes. In J. M. Atkinson and J. Heritage, editors, *Structure of Social Action: Studies in Conversation Analysis*. Cambridge University Press.

Matt Richardson and Pedro Domingos. 2006. Markov logic networks. *Machine Learning*, 62(1-2):107–136.

Craige Roberts. 1996. Information structure: Towards an integrated formal theory of pragmatics. In Jae Hak Yoon and Andreas Kathol, editors, *OSU Working Papers in Linguistics*, volume 49: Papers in Semantics, pages 91–136. The Ohio State University Department of Linguistics, Columbus, OH. Revised 1998.

Robert van Rooy. 2003. Questioning to resolve decision problems. *Linguistics and Philosophy*, 26(6):727–763.

Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53:361–382.

Anna-Brita Stenström. 1984. Questions and responses in English conversation. In Claes Schaar and Jan Svartvik, editors, *Lund Studies in English 68*, Malmö Sweden. CWK Gleerup.

Henk Zeevat. 1994. Questions and exhaustivity in update semantics. In Harry Bunt, Reinhard Muskens, and Gerrit Rentier, editors, *Proceedings of the International Workshop on Computational Semantics*, pages 211–221. ITK, Tilburg.