Squib 2: Applying the Iterated RSA model to implicatures

Hanson Lu

LING 23200, Topics in Semantics and Pragmatics, Spring 2019

1 Introduction and Background

Developing on the Rational Speech-Act (RSA) model first proposed in literature such as Goodman and Stuhlmüller (2013), Bergen et al. (2016) expands the applicability of the RSA model to M-implicatures (Horn, 1986) and embedded implicatures (Chierchia et al., 2012).

As cited from Bergen et al. (2016), the following examples (2) and (3a) are instances where M-implicatures and embedded implicatures can be derived respectively.

M-implicatures are, simply put, an association between a more complex (or more marked) utterance with a meaning that would expected to occur more rarely. For instance, (2b) "will usually be interpreted to mean that John will not finish the homework," while (2a) does not have this implicature. The type of embedded implicatures discussed in Bergen et al. (2016), which are to be investigated in this paper, are ones where a weaker expression (such as *or*) is embedded in a downward entailing environment. Without the downward entailing environment, for example, in (3b), a strengthened implicature of "John didn't talk to both but talked to one of them", is expected, in other words, *or* is strengthened to *exclusive or* (XOR in short). On the other hand, such strengthening is strongly is strongly dispreferred in the case of (3a), which if it is strengthened, we would get an erroneous meaning.

(1) a. Kai ate some of the cookies.

b. Kai ate some but not all of the cookies.

(2) a. John can finish the homework.

b. John has the ability to finish the homework.

(3) a. John didn't talk to Mary or Sue.

b. John talked to Mary or Sue.

Bergen et al. (2016) shows that the original RSA model proposed in Goodman and Stuhlmüller (2013) was only able to derive basic scalar implicatures, such as (1a) implicating (1b), but not able to derive more complex ones, including M-implicatures and embedded implicatures. They then propose a modification to the original RSA model that enables these implicatures to be derived. The modification is basically incorporating an uncertainty in "lexicon". A "lexicon", as defined by Bergen et al. (2016), is in general terms an assignment of truth values to utterances. Their formulation successfully derives M-implicatures such as (2), and also predicts that strengthening is dispreferred in embedded implicatures as in (3a). They also leave the open question that in cases where the weaker scalar expression is emphasized, as in (4), the strengthened XOR meaning of *or* is indeed derived, and whether this could be modeled by the RSA framework.

(4) John didn't talk to Mary OR Sue, he talked to both.

In this paper, I would like to apply an alternative RSA model, namely the *Iterated RSA Model*, which is similar to the basic RSA model except that the pragmatic reasoning is performed at the level of each word, rather than a whole utterance. This model is proposed in Cohn-Gordon et al. (2019), and is presented below. My goal is to apply this model to the case of simple scalar implicatures and the specific case of embedded implicatures as in (3a). I show that the model is able to successfully derive the strengthened meaning of *some* to $\exists\neg\forall$, and avoid strengthening the *or* when embedded under the scope of negation.

2 The model

Similar to the standard RSA model, the Incremental RSA model mainly works by a recursive reasoning between the speaker and listener. In the following I will give an outline of how the model is defined, according to Cohn-Gordon et al. (2019). The code that I wrote for this model may be accessed at `https://github.com/hansonhl/IteratedRSA`.

2.1 Setup of iterated RSA model

For the setup of the model, we are given a set of possible complete utterances U and a set of world states W. Because we are reasoning incrementally word by word, we need to know when a sentence begins and ends. Hence in our model at the beginning and end of each complete utterance we attach the tags `<start>`

and `<end>`, which are treated normally as other words. We are also given a function that takes care of the semantics of words, called an "interpretation function" (in Cohn-Gordon's terms), or effectively the "lexicon" function in Bergen et al. (2016). For a complete utterance $u \in U$ and a world state $w \in W$, this function is defined as:

$$\mathcal{L}(u, w) = [\![u]\!](w) = \begin{cases} 1 & \text{if the utterance } u \text{ is true in the world state } w \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where $\mathcal{L}(u, w)$ and $[\![u]\!]w$ are just alternative notations. I will use the latter notation for coherence.

Because the iterated RSA model reasons with words one at a time, its reasoning takes into account of *incomplete* utterances. For any string of words c, let $U[c] \subseteq U$ be the subset of possible complete utterances such that

$$U[c] = \{u \in U : c \text{ is a prefix of } u\} \tag{2}$$

Then we define the "semantics" of an incomplete utterance c as

$$[\![c]\!](w) = \frac{\text{Number of utterances } u \in U[c] \text{ such that } [\![u]\!](w) = 1}{\text{Total number of utterances in } U[c]} \tag{3}$$

Using the above setup, we define the literal listener L_0 is defined as:

$$L_0(w \mid c, word) \propto [\![c + word]\!](w) \tag{4}$$

Here, L_0 gives the probability of a world state w, given an an incomplete sentence c and a newly heard word $word$. $c + word$ is the concatenation of c and $word$[1]. For the sake of simplicity, different from Bergen et al. (2016)'s formulation, we assume all prior possibilities for worlds states $P(w)$ to be equal. Hence we do not include the term $P(w)$ in the equation.

Then we define the pragmatic speakers and listeners recursively. For $n \geq 1$:

$$S_n(word \mid c, w) \propto e^{\log L_{n-1}(w|c,word) - cost(word)} \tag{5}$$

$$L_n(w \mid c, word) \propto S_n(word \mid c, w) \tag{6}$$

[1] If $word$ is the `<end>` symbol, then $c + word$ forms a complete utterance, and we use the definition of $[\![\cdot]\!]$ in (1).

Here S_n gives a probability for each possible word to say, given a world state w and incomplete sentence c, reasoning on the listener L_{n-1}. The set of possible words consists of all the words that appear in the set of utterances U. For simplicity we let the cost function take the value of 1 for each word.

2.2 Speaker and listener models that reason on complete utterances based on the iterated version

For each level n of recursion, we define the probability of a complete utterance u given a world state as

$$S_n^{UTT}(u \mid w) = \prod_{i=1}^{l} S_n(u_i \mid c = (u_1 + u_2 + ... + u_{i-1}), w) \tag{7}$$

Here l is the length of the complete utterance u, u_i is the ith word in u, and $(u_1 + u_2 + ... + u_{i-1})$ is the concatenation of words in positions 1 to $i-1$. In practice, according to Cohn-Gordon et al. (2019), if we are given a world state w and we would like to find the most probable utterance whose probability is defined using (7), we may perform "*greedy unrolling*", where we start with the tag `<start>`, and we choose a word that maximizes the probability $S_n(word \mid c = \texttt{<start>}, w)$. Let this word be u_2. Then we append u_2 to c, and proceed on to the next position and choose the word that maximizes the probability $S_n(word \mid c = \texttt{<start>} + u_2, w)$, and append that word to c etc., until the most probable word is the ending tag `<end>`. We are guaranteed that the words we choose in this manner gives us the most probable complete utterance.

Corresponding to (7), I also propose two possible methods for a pragmatic listener reasoning about a world state, given a *complete* utterance u. The first method cam be expressed using the following equations.

$$W(u,i) = \underset{w}{\operatorname{argmax}}(L_n(w \mid c = (u_1 + ... + u_{i-1}), u_i)) \tag{8}$$

$$P(u,i) = L_n(W(u,i) \mid c = (u_1 + ... + u_{i-1}), u_i) \tag{9}$$

$$i^* = \underset{i \in 1,...,|u|}{\operatorname{argmax}} P(u,i) \tag{10}$$

$$L_n^{MAX}(u) = W(u, i^*) \tag{11}$$

Here, for each position i in the sentence u, where i ranges from 1 to the length of u (denoted by $|u|$), I first get the world with the greatest probability calculated by the rational speaker, given the incomplete utterance consisting of previous words $c = (u_1 + ... + u_{i-1})$, and the word at the current position u_i. I denote this most possible world inferenced at position i of utterance u as $W(u,i)$, and is expressed by equation (8). Then I let the probability of that world be $P(u,i)$, as shown by equation(9). Finally, I get the position i^* that maximizes $P(u,i)$ (equation (10)), and take the world at that position as the world that the pragmatic listener inferences for the entire utterance u. In the end L_n^{MAX} does not give a probability, but directly gives the world state that turned out to be most probable at some point during the utterance of u.

The second method is expressed by equation (12), and gives a probability for each world state w based on the rational speaker. Again, we assume that all world states have equal prior probability.

$$L_n^{UTT}(w \mid u) \propto S_n^{UTT}(u \mid w) \tag{12}$$

I implemented this model using python script, which is accessible at ....

3 Simple scalar implicature

As an example, I will first apply the model to the simple case of scalar implicature involving the utterances *some*, *all* and the set of world states $\{\forall, \exists\neg\forall\}$. The function $[\![u]\!](w)$ is defined as:

$$[\![\text{all}]\!](\forall) = 1, [\![\text{all}]\!](\exists\neg\forall) = 0, [\![\text{some}]\!](\forall) = 1, [\![\text{some}]\!](\exists\neg\forall) = 1$$

Note that here I treat the single words "some" and "all" as complete utterances, instead of using fully-fledged sentences like (1a). This is because even if we used them, the only difference among the possible utterances would be one word, with other parts remaining the same, and other calculations would turn out to be the same as well. Hence I make the above simplification.

The probabilities given by the speaker S_1 are shown in the following tree format, similar to those in Cohn-Gordon et al. (2019). The number on each edge gives the probability of choosing a word given previous words, for instance $S_1(\textit{all} \mid c = \texttt{<start>}, w = \exists\neg\forall) = 0.0$.

Hence from this figure we may obtain the most probable sentence "`<start>` some `<end>`". Using the "greedy unrolling" method, we first decide between "some" or "all" after `<start>`. We choose "some" because it has the higher probability. Our next step is obviously to choose `<end>`, as other words are assigned the probability zero (not shown here). We may further look at the results for S_0 and L_0 when

	$\exists\neg\forall$	$\forall$
some	0.5	0.5
all	0.0	1.0

(a) $L_0(w = * \mid c = \texttt{<start>}, \mathit{word} = *)$

	some	all	<start>	<end>
$\exists\neg\forall$	1.0	0.0	0.0	0.0
$\forall$	0.3333	0.6667	0.0	0.0

(b) $S_1(\mathit{word} = * \mid c = \texttt{<start>}, w = *)$

Table 1: Probabilities when reasoning what word to say after <start>

Utterance	some	<end>
Most probable world	$\exists\neg\forall$	$\exists\neg\forall$
Probability	**0.75**	0.5

(a) Results for $L_1^{MAX} = \exists\neg\forall$

World state	$\exists\neg\forall$	$\forall$
Probability	0.75	0.25

(b) Results for L_1^{UTT}

Table 2: Results for complete listener models when hearing the utterance '<start> some <end>'

reasoning the first word after start, which are shown in table 1a and 1b. The results are similar to that of the baseline model, proving that the effectiveness of the iterated speaker model.

The results of the two complete utterance listener models (defined in equations (8) to (12)) are shown in table 2. In table 2a, each cell in the row "Most probable world" is the most probable world when the listener reasons based on the word shown at the top of the column and all words preceding it forming an incomplete utterance. This corresponds to $W(u, i)$ defined in (8). The highest probability among all positions is then shown in bold, and the world state in that column is taken to be the result. We can see from these results that the listener in both listener models does indeed associate "some" with the world state $\exists\neg\forall$, which is desired.

4 Embedded *or* under the scope of negation

In this section I will apply the iterated RSA model to the case of embedded implicature discussed in Bergen et al. (2016) where the disjunctive operator *or* is embedded under the scope of negation. First, I use the same setup with regards to possible complete utterances, and world states, which are shown in table 3. The world states $\varnothing$, M, S, MS represent the world state where no one, only Mary, only Sue, and both of them respectively are, for example, invited to a party, being talked to by John, etc. The leftmost column gives all

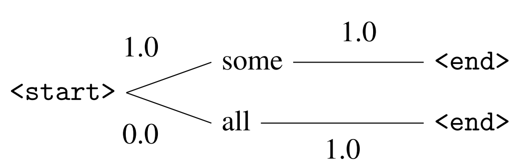

Figure 1: Incremental reasoning for the world state $\exists\neg\forall$

$[\![u]\!](w)$	$\varnothing$	M	S	MS
mary	0	1	0	1
sue	0	0	1	1
not mary	1	0	1	0
not sue	1	1	0	0
mary and sue	0	0	0	1
sue and mary	0	0	0	1
mary or sue	0	1	1	1
sue or mary	0	1	1	1
not mary and sue	1	1	1	0
not sue and mary	1	1	1	0
not mary or sue	1	0	0	0
not sue or mary	1	0	0	0

Table 3: World states, possible utterances, and values used for $[\![u]\!](w)$

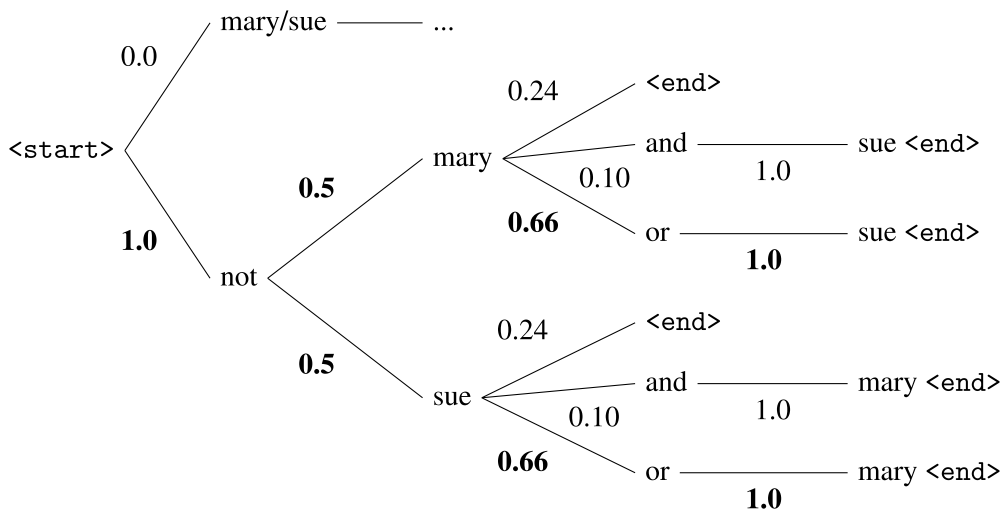

Figure 2: Iterated reasoning of the 2nd level pragmatic speaker given the world state $\varnothing$

possible utterances using the grammar (which is not replicated here) that Bergen et al. (2016) sets up for this scenario. For the binary operators *and* and *or*, I included the case where the operands may be reversed. Similar to the case with scalar implicatures, I do not take into the account of the entire sentence "John talked to Mary and Sue" but only extract the part that is logically relevant.

Running the iterated RSA model, I get the following tree of probabilities for the pragmatic speaker S_2, when given the $\varnothing$ word state, shown in figure 2. Note that the branch to "mary" or "sue" after `<start>` is not shown here, as our model assigns both options 0 probability. Other ungrammatical situations (situations that would result in an utterance that is not among the above list) such as `<end>` following "not" are assigned 0 probability automatically and are not shown as well.

	$\varnothing$	M	S	MS
or	1.0	0.0	0.0	0.0
and	0.33	0.33	0.33	0.0
<end>	0.50	0.0	0.50	0.0

(a) $L_0(w = * \mid c =$ `<start>` not mary$, word = *)$

	or	and	<end>
$\varnothing$	0.55	0.18	0.27
M	0.00	1.00	0.00
S	0.00	0.40	0.60
MS	0.33	0.33	0.33

(b) $S_1(word = * \mid c =$ `<start>` not mary$, w = *)$

Table 4: Probabilities when reasoning what word to say after <start>

From this figure, the two most probable sentences are "<start> not mary or sue <end> " and "<start> not sue or mary <end>", which is what is desired. Both are equally probable, reflecting the symmetric nature of *or*. One crucial step in generating this result is the reasoning of the speaker with the incomplete utterance "<start> not mary" and deciding what to say next among <end>, "and", "or". We show the probabilities relevant to this step in table 4. We can observe that when in the world state $\varnothing$, "or" at this stage will be the most informative for the literal listener L_0. Hence S_1 will prefer to choose "or". This probability is increased in S_2, based on the reasoning of L_1. The same reasoning applies when the incomplete sentence is "<start> not sue".

We may also look at the results of complete utterance listener models when reasoning on the sentence "<start> not mary or sue <end> ", shown below in table 5. We see that both listener models correctly predict $\varnothing$ as the world state, which is desired. An interesting result is that after hearing "not mary", the most probable world state that the listener infers is S, though its probability is low. This would make sense intuitively, as for instance if we were asked the question "Mary and Sue were at the party. Who did John talk to?", and we reply "Not Mary." and paused there, then our answer would be interpreted as John actually talked to Sue. However, we chose that answer instead of directly answering the short affirmative answer "Sue." This may have some implications regarding M-implicatures and ignorance implicatures which are not yet fully captured by our model, for instance this may imply that I am certain that John didn't talk to Mary, but uncertain about Sue, even though to some extent this is reflected by the low probability.

Utterance	not	mary	or	sue	<end>
Most probable world	$\varnothing$	S	$\varnothing$	$\varnothing$	$\varnothing$
Probability	0.6	0.33	**0.62**	0.25	0.25

(a) Results for $L_1^{MAX} = \varnothing$

World state	$\varnothing$	M	S	MS
Probability	1.0	0.00	0.00	0.00

(b) Results for L_1^{UTT}

Table 5: Results for complete listener models when hearing the utterance '<start> some <end>'

5 Conclusion

In conclusion, we have shown that the iterated RSA model can successfully derive the strengthened meaning of *some* to $\exists\neg\forall$, and avoid strengthening the *or* when embedded under the scope of negation. I have also provided two methods that model the pragmatic listener's response to a whole sentence, given an iterated speaker model.

We yet need to consider other forms of implicatures such as ignorance implicatures and M implicatures. In order to successfully derive M implicatures, it may be necessary to revise the definition of the cost function in the iterated model, as currently each word is treated with an equal cost, thus having no effect on the outcome. One way may be to factor in the cost of possible continuations for the cost of a word. To deal with ignorance implicatures, we need to separate worlds states from observations, as Bergen et al. (2016) did. Moreover, we have not yet solved the problem of stressed *or* under the scope of negation actually being strengthened. To deal with this problem we may consider revising the semantics function $[\![c]\!](w)$ to consider the counts of *incomplete* sentences that are grammatical and part of a complete utterance, for instance, extracting the incomplete sentence "Mary or Sue" from "not Mary or Sue", and only reasoning on the former.

References

Leon Bergen, Roger Levy, and Noah Goodman. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics*, 9:1–83, 2016.

Gennaro Chierchia, Danny Fox, and Benjamin Spector. Scalar implicature as a grammatical phenomenon. In Claudia Maienborn, Klaus von Heusinger, and Paul Portner, editors, *Semantics: An international handbook of natural language meaning*, volume 3. de Gruyter, 2012.

Reuben Cohn-Gordon, Noah Goodman, and Christopher Potts. An incremental iterated response model of pragmatics. In *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, pages 81–90, 2019. doi: 10.7275/cprc-8x17. URL `https://www.aclweb.org/anthology/W19-0109`.

Noah Goodman and Andreas Stuhlmüller. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science*, 5:173–184, 2013.

Laurence Horn. Towards a new taxonomy for pragmatic inference: Q-based and r-based implicature. In Deborah Schiffren, editor, *Meaning, form and use in context*, pages 11–42. Washington DC: Georgetown University Press, 1986.