Discourse Understanding and World Knowledge

MEGUMI KAMEYAMA*

Discourse processing is a sequence of dynamic operations on transient information states. As discourse unfolds, the participants keep track of a shared information state that is dynamically updated in a chain of transitions. This shared information state is a bundle of both explicitly asserted and implicitly assumed pieces of information. The role of a linguistic utterance is to make explicit only what does not follow from the currently shared information. Discourse understanding thus combines both linguistic interpretation and commonsense reasoning. It is essential that a theory of discourse account for their interactions.

These points are illustrated with an analysis of a piece of discourse—the AI puzzle of 'Missionaries and Cannibals'. The approach is an information-based dynamic discourse analysis with the following features: (1) an information state consists of salient situation types and information pieces, (2) the use of each referring expression type has a set of dynamic constraints in terms of pre- and post-conditions on the information state. The approach exposes the systematic dynamicity in linguistic processing that has tended to be either taken for granted without explanation or neglected in the studies of commonsense reasoning in AI. The same approach is used to analyze equivalent texts in English and Japanese demonstrating its cross-linguistic applicability.

1. Introduction

Natural language as a communication medium is an intricate system of underspecification. It is designed to work together with the human reasoning capability based on the knowledge of what goes on in the normal world. It is used with the general principle of *explicitly mention only what does not follow from the current shared knowedge of the discourse participants.* Linguistic communication leaves a lot of crucial information implicit as part of the shared assumptions. Discourse understanding is a combination of linguistic interpretation and commonsense reasoning. In addition to the difficulty in formalizing each of these alone, a theory of discourse requires a formal account of how they interact at each step of discourse progression.

I take a dynamic view of discourse emerging from the current developments in computational discourse analysis and logical pragmatics. It is the view that each utterance in discourse serves to update the current information state shared among the discourse participants. This updating can take the form of incremental addition or revision. It can also be a suspension of the current information state altogether to open a brand-new information state or to resume some other previously suspended state. A new shared information state results from processing (i.e., either generating or understanding) each utterance, and this new state in turn serves as the context in which the next utterance is processed.

This dynamically maintained *information state* crucial to all discourse operations in this view consists of both explicitly asserted pieces and implicitly assumed pieces of information. As a first approximation, we may say that language gives rise to the explicit, and the shared world knowledge gives rise to the implicit. Language alone, however, does not fully specify the explicit information that an utterance is intended to contribute to the shared information state.

I wil illustrate these general points with an analysis of a piece of discourse, the well-known Al puzzle called "missionaries and cannibals".

2. Missionaries and Cannibals

Three missionaries and three cannibals come to a river. A rowboat that seats two is available. If the cannibals ever outnumber the missionaries on either bank of the river, the missionaries will be eaten. How shall they cross the river? (from McCarthy 1980)

Upon reading this puzzle, the puzzler is expected to come up with a strategy of rowing the boat back and forth that gets them all across the river without having any of the missionaries eaten by the cannibals. For solving this puzzle, it is crucial to keep track of the numbers of missionaries and cannibals throughout the course of events that constitute the solution.

One representation scheme for the solution, commonly regarded as optimal, is that of Amarel's (1971). It is a

*SRI International Al Center, 333 Ravenswood Ave., Menlo Park, CA 94025 and The Center for the Study of Language and Information, Stanford University.

state space representation where each state is a triple with the numbers of (i) missionaries, (ii) cannibals, and (iii) boats on the starting bank of the river. The initial state is 331, the desired final state is 000, and one solution is given by the sequence (331, 220, 321, 300, 311, 110, 221, 020, 031, 010, 021, 000).

3. Implicit Assumptions

McCarthy (1980) discussed the reasoning that goes from the English statement of the problem to Amarel's state space representation. He skipped the linguistic analysis of the English text, and started from what he thought was a direct translation of the English statement into first order logic. Even with this simplification, he found that Amarel's representation was not an ordinary logical consequence of these sentences. He gave two reasons:

(1) Essential facts are not stated if they follow from commonsense knowlesge. For instance, nothing is said about the properties of boats such as the fact that rowing across the river does not change the numbers of missionaries or cannibals or the capacity of the boat, or it is not stated that situations change as a result of action.

(2) No relevant objects exist in certain categories except those whose existence follows from the statement of the problem and commonsense knowledge. For instance, since no bridge is mentioned in the statement of the problem, there is no bridge, and since nothing is said about the abnormality of the boat, it has all the normal properties such as having oars and it does not have a leak.

Based on (1), McCarthy proposed that what follows from commonsense knowledge, at least the relevant parts of it, be also expressed in first order logic and be added to the statement of the problem. Based on (2), he identified the need for nonmonotonic reasoning. He proposed circumscription as one candidate for accomplishing it, and concluded that the general notion of "something abnormal about x" should be part of commonsense ontology (i.e., the thing that exists).

Following McCarthy, I assume that the shared information state of evolving discourse has access to relevant commonsense assumptions. I also assume that the reasoning involved is nonmonotonic, that is, unmentioned objects and events are implicitly assumed to be nonexistent, and objects and events mentioned are assumed to be "normal" unless otherwise indicated.

4. Dynamic Discourse Analysis

Whatever is explicitly said in language should directly be part of the shared information state. How much does language really make explicit, then? To examine this question, I will now analyze the English text of the puzzle to reveal precisely the part of explicit information that language alone contributes.

4.1 Chain of Transitions

The English text consists of four sentences. Let us assume that these are the utterance units u1, . . . , u4 that trigger updating of the shared information state in a sequence:

---

u1. Three missionaries and three cannibals come to a river.
u2. A rowboat that seats two is available.
u3. If the cannibals ever outnumber the missionaries on either bank of the river, the missionaries will be eaten.
u4. How shall they cross the river?

We assume that there is an initially shared information state i0 at the onset of discourse, which sets certain relavant discourse parameters. Then as a result of processing u1, i0 is updated and a new information state i1 is created with the addition of the facts expressed in u1. The next utterance u2 will then be processed with respect to i1, and so on. In general, under this dynamic view, discourse is a chain of transitions such that

i_n (the current information state)
$+u_n$ (the next utterance)
$=i_{n+1}$ (the next information state)

We can also see it this way: Each utterance is an "action" that triggers a transition from the current information state to the next. (Van Benthem 1991 gives logical possibilities for modelling such information states and a set of transition relations on them.) This is analogous to the common AI view of a non-linguistic action (such as climbing a tree) as defining a transition from one situation to the next. (as in McCarthy and Hayes's situation calculus.) The important difference is that the transient information state in linguistic actions is a mental object. It is a representation of the information that the discourse participants believe to share among them. They do not need to "believe" what is in this shared representation. They only have to "agree" on it, or "believe to agree" on it. This view of discourse shares its essence with a number of existing insights in computational discourse analysis and logical pragmatics, e.g., scorekeeping (Lewis 1979), Discourse Representation Theory (Kamp 1981), File Change Semantics (Heim 1982), language games (Hintikka and Kulas 1983), the attentional state and intentional structure (Grosz and Sidner 1986), discourse grammar (Scha and Polanyi 1988), Dynamic Predicate Logic (Groenendijk and Stokhof 1988), epistemic state dynamics (Gärdenfors 1988), and the conversational record (Thomason 1990).

4.2 Representing the Information State

This shared information state is a complex bundle of both explicitly asserted and implicitly assumed pieces of information. I will first summarize my view on what goes into the implicit part, then turn to an extensive discussion of the focus of this paper the explicit part that results from linguistic descriptions.

The implicit part can be divided into the *background knowledge* and the shared *implicit assumptions*. The background knowledge includes the world knowledge and the knowledge of the social and cultural conventions that the discourse participants share. This is an open-ended mass of information—a huge iceberg whose tip is revealed in a discourse. The shared implicit assumptions include the discourse participants' goals, elements in the current utterance situation, and assertions generated by *conversational implicature*. Conversational implicature is a general phenomenon whereby a speaker A, upon explicitly asserting p, also implicitly asserts q by making the hearer recognize A's intention to assert q (Grice 1957). For instance, an action of looking at one's watch could imply his desire to leave. This sort of implicature is generally cancellable by an explicit denial, adding further complications. Correct recognition of implicature is thus essential for keeping track of the correct information state. The aspect of commonsense reasoning that McCarthy noted above would largely fall under conversational implicature.

My specific choice in representing the shared information state draws on the notion of situation types (Barwise and Perry 1983) that support salient facts. Four situation types combine to characterize an utterance: Described Situations, Utterance Situations, Discourse Situations, and Phrasal Situations. *Described situations* (S) are where information pieces are made factual. Situations described in discourse need not be "real" because people can talk about any imaginary or hypothetical situations, and the "truth" as well as "existence" are entirely situation-dependent. A number of different described situations typically arise even in a coherent discourse. The *utterance situation* (US) is the indexical context of the utterance. It holds crucial information about the physical context of the utterance, e.g., who the speaker is, who the addressee is, where it is uttered, when it is uttered, etc. Entities in the uteracne situation are accessed with indexical or deictic references. *Discourse situations* (DisS) (Gawron and Peters 1990, Kameyama et al. 1991) correspond to the transient information states under the present dynamic view of discourse. The *phrasal situation* represents the surface form of an utterance.

When a piece of information (or *infon* in Devlin (1991)) f is a fact in a described situation s, s *supports* f (notated below as s $\models$ f). We can see such information pieces as equivalent to *discourse entities* in computational discourse analysis. Discourse entities are *evoked* by explicit linguistic mentions of objects and events, and *accessed* by referring expressions in subsequent discourse (Webber 1983). These are "cognitive elements" (Grosz 1977, Sidner 1983) that mediate the objects/events in actual worlds and people's concepts about them. They are the building blocks for the information content that an utterance contributes, and hence, for the shared information state in an evolving discourse. I will hencefouth use the term *information pieces* in place of discourse entities.

In the following analysis, I will distinguish between

event-type information pieces (e1, . . . , e*N*)

and

object-type information pieces (n1, . . . , n*N*)

Only the former are directly associated with temporal intervals. Intuitively, they represent the events, states, and processes described by tensed verb phrases, non-tensed verb phrases (e.g., *to understand it, understanding it*), nominal phrases morphologically derived from verbs (e.g., *destruction, maintenance*), and *event nouns* (e.g., *trip, party, picnic*). As part of discourse understanding, the temporal intervals mentioned in discourse must be incorporated into a coherent, but not necessary fully specified, temporal network.

When something is newly mentioned and mutually understood, it goes into the shared information state and can be further talked about in the subsequent discourse. This is done in the present analysis by entering into the information state a set of described situations, information pieces that they support (i.e., facts), and accessible entities (i.e., sets of individual objects) in the domain of discourse. They are evoked by the following rules:

Evoking information pieces

Information pieces arise by projection of the syntactic-semantic structure of an utterance. All the referring expressions—associated with the lexical N(ominal) and V(erbal) "heads" in an utterance—evoke distinct information pieces. The hierarchical structure of an utterance gives rise to the hierarchical structure among corresponding information pieces.

Evoking entities in the domain of discourse

Information pieces are "about" certain (sets of) objects and events. Such *entities* in the current domain of discourse are accessible and can be further talked about. These elements are partially ordered with respect to their relative *salience* representing the attentional aspect of discourse (Grosz and Sidner 1986).

In theory, relations among temporal intervals could be fully specified using, e.g., Allen's (1983) 13 temporal relations, but events described in discourse are not always locatable in a fully specified temporal network. What we get from the interpretation rules of tense and aspect plus commonsense is typically a network of events related with one another in underspecified temporal relations that the grammar of tense and aspect in the language minimally distinguish. It seems that more specific temporal relations are explicitly given only as needed for the purpose of the given discourse. Spatial relations in discourse are much the same way, that is, they are specified only as much as appropriate for the discourse purpose. I will not discuss spatial references in this paper.

Evoking described situations

An utterance evokes one or more described situations. For instance, the utterance *John thought Mary would come* evokes two described situations, one where John's thinking took place some time prior to the time the speaker utters it (call it s1), and the other within this thinking of John's (call it s2) where Mary's coming at some future point in time was a fact. It is clear that s2 is in some sense dependent on s1, but the two situations can support totally independent sets of facts. Each new described situation that arises from an utterance must be merged into the current information state either by being unified with an existing situation or by being related to one. As part of discourse understanding, the described situations in a given discourse must form a coherent, but not necessary fully specified, situation network.

Evoking utterance situations

Each utterance evokes at least one new utterance situation. In reportive speech such as *Max said "Felix is at school"*, and embedded utterance situation is evoked with Max as the speaker.

4.3 Dynamic Resolution of Referring Expressions

The context-dependence of various referring expressions is dynamically characterized in this approach. Each referring expression type is given a set of disjunctive pairs of a pre-condition and a post-condition on the "context". The "context" here can be the intra-sentential linguistic context, the current shared information state (discourse situation), or the current utterance situation. A *pre-condition* imposes a certain requirement on the current context concerning the (non-)existence and salience of entity x specified by the expression. when entity x is in the information state i, x is either in the currently salient situation s, or x "corresponds to" x′ in a situation s′ related to s. If the pre-condition is met, the context is updated to meet the *post-condition*. We can also allow pre-conditions to impose changes in the current shared context (in the pre-updating state) as a case of "accomodation" (Thomason 1990). This dynamic characterization of referring expression types is similar to how programs are seen as predicate transformers in computer science. It is also similar to Heim's (1982) approach to (in)definiteness. The operations on pre- and post-conditions here can be implemented as "assumption discharging" in Pereira and Pollack's forthcoming incremental interpretation approach.

The following basic conditions for English referring expressions suffice for the present purpose:

Given the current shared information state i and the next information state i+, when a referring expression r in the current utterance u has the following type, the following pre- and post-conditions are imposed on the entity x that r specifies: (Read "in i&" below as "in i or mentioned in u".)

Conditions on English Referring Expression Types (Part)

1. INDEFINITE

Case 1: pre-condition: x is not in i&
post-condition: x is in i+

Case 2: pre-condition: x is a subset of a group entity g in i&
post-condition: x is in i+

2. DEFINITE

Case 1: pre-condition: x is in i&
post-condition: x is in i+

Case 2: pre-condition: x is not in i, but accessible in the background shared information state as a unique entity (e.g. "the moon")
post-condition: x is in i+

3. 3RD PERSON PRONOMINAL

Case 1: pre-condition: x is in i& and is salient
post-condition: x is in i+ and is salient

Case 2: pre-condition: x is in the current US (deixis)
post-condition: x is in i+ and is salient

4. EITHER

Case 1: pre-condition: there are exactly two members in x, x is in i&
post-condition: x is in i+

4.4 Resolution of Tense

Tense interpretation and temporal reference resolution can also receive a dynamic characterization analogous to that of referring expressions. The anaphoric property of tense in discourse (Partee 1984) is commonly characterized using Reichenbach's (1947) *reference time*—the temporal vantage point. I will assume this line of approach here to say that each information state has one temporal interval, say, rt, as its current reference time. Then, in general, we can say that rt is either kept or shifted to another time interval in i+. That is, rt and rt+ are either the same or different. Without going into more detail, which is discussed in Kameyama. et al. (in preparation), I will give only a few illustrative conditions here:

Given a tense operator (e.g., PAST, PRES, FUTURE) with a time interval t as its argument, in an utterance u describing facts in situation s, with the current reference time rt of the current information state i:

Conditions on English Tenses (Preliminary)

1. PRES

pre-condition: t coincides with the time of the current US, the temporal span of s includes t
post-condition: rt+ is either the same or different from rt

2. PAST

pre-condition: t precedes the time of the current US, the temporal span of s includest t post-condition: rt+ is either the same or different from rt

3. FUTURE

pre-condition: t follows the time of the current US, the temporal span of s includes t

post-condition: rt+ is either the same or different from rt

4.5 Updating the Shared Information

A global thread of discourse associated with the general discourse structure can also be localized under a uniform dynamism. Given the current shared information state i and the current utterance u, we first evoke a new described situation s that supports what is described by u. Then a general updating operation (add, i, s) breaks down into a number of specific 'add' operations. In each case, the new information obtained from u is added to the information already in the same described situation (with possible revisions). The preferred operation is to RETAIN the currently salient situation, say, s!, and s=s! (Cases 1–2 below). Otherwise, the transition can be one of EMBED, SUSPEND, and RESUME (Cases 3–5 below):

Information State Updating Conditions (Preliminary)

(ADD, i, s)

Case 1: pre-condition: i contains no situations
post-condition: s is in i+

Case 2: pre-condition: s! is salient in i
post-condition: s=s!, s! is salient in i+

Case 3: pre-condition: s! is salient in i
post-condition: s is in a subordinating relation to s!, s is salient in i+

Case 4: pre-condition: s! is salient in i, a discourse interruption at u
post-condition: supend s!, s is salient in i+

Case 5: pre-condition: s! is salient in i, a discourse resumption at u, there is s′ either suspended or superordinate to s!
post-condition: s=s′, s′ is salient in i+

5. Analysis of the English Text

In the analyses below, I will use the following notations: Variables, or restricted parameters, are notated with lower case letters. Unrestricted parameters are in upper case letters with question marks, e.g., Q?. They correspond to uninstantiated objects and facts. Object-type information pieces are notated as $n_1, \ldots, n_n$, and each is defined as a conjoined set of restrictions on a parametric object. Event-type information pieces are notated as $e_1, \ldots, e_n$. Each is an n-place relation over parametric objects, the first of which is the Davidsonian 'event' variable followed by N-2 arguments, ended by a temporal interval notated as $t_1, \ldots, t_n$. Information pieces can also be composed from one or more discourse entities with n-ary relations in the form (relation, $\text{entity}_1, \ldots, \text{entity}_n$). Situations are notated as $s_1, \ldots, s_n$. When a situation supports a discourse entity x, or x is a fact in s, it is notated as $s \models x$.

Brackets [ ] indicate the linguistic units that evoke distinct information pieces.

See, e.g., Grosz and Sidner (1986), Polanyi (1988), Hobbs (1990) and references cited there for examples and discussions of these discourse structure operations.

initial information state DisS0:

Facts: { }
Entities in the Domain: { }

US1: [[[Three missionaries] and [three cannibals]] come to [a river]].

S1⊨{ n1=(join,{m,c},mc),
n2=(and,(missionary,m),(3,m)),
n3=(and,(cannibal,c),(3,c)),
n4=(and,(river,r),(1,r)),
e1=(and,(come,p1,mc,r,t1),(pres,t1))}

Condition to be removed—>	Result of remval:
(add,DisS0,S1)	—>S1 is in DisS1. [Case1]
(indef,m)	—>m is in DisS1. [Case1]
(indef,c)	—>c is in DisS1. [Case 1]
(indef,r)	—>r is in DisS1. [Case1]
(pres,t1)	—>t1 coincides with the time of US1

information state DisS1:

Facts: {S1⊨{n1,n2,n3,n4,e1}}
Entities in the Domain: {mc,m,c,r,p1,t1}

US2: [[A rowboat that seats two] is available].

S2⊨{ n5=(and,(rowboat,b),(1,b),(seat_of,b,2))
e2=(and,(available,p2,b,t3),(pres,t2))}

Condition to be removed—>	Result of removal:
(add,DisS1,S2)	—>S1=S2. [Case2]
(indef,b)	—>b is in DisS2. [Case1]
(pres,t2)	—>t2 coincides with the time of US2.

information state DisS2:

Facts: {S1⊨{n1,n2,n3,n4,e1,n5,e2}}
Entities in the Domain: {mc,m,c,r,p1,t1,b,p2,t2}

US3: If [[the cannibals] ever outnumber [the missionaries] on [either bank of [the river]]], [[the missionaries] will be eaten].

S3 ⊨ { n6=(and,(cannibal,c2),(plural,c2)),
n7=(and,(missionary,m2),(plural,m2)),
n8=(and,(bank,a),(either,a),(of,a,r2)),
n9=(and,(river,r2),(1,r2),
e3=(and,(outnumber,p3,c2,m2,t3),
(location,p3,a),(ever,t3),(pres,t3)),
n10=(and,(missionary,m3),(plural,m3)),
e4=(and,(be_eaten,p4,m3,t4),(future,t4)),
e5=(if,e3,e4)}

Condition to be removed—> Result of removal:

(add,DisS2,S3)	—>(hypothetical_on,S3,S1) [Case3]
(def,c2)	—>c2 % c. [Case1] (c2 corresponds to but does not resolve to c.)
(def,m2)	—>m2 % m. [Case1]
(def,r2)	—>r2 % r. [Case1]
(either,a)	—>a normal river has exactly two banks [Case1]
(def,m3)	—>m3=m2. (m3 resolves to m2.) [Case1]
(pres,t3)	—>t3 is sometime after the time of US3.
(future,t4)	—>t4 is sometime after t3.

information state DisS3:

Facts: {S1 ⊨ {n1,n2,n3,n4,e1,n5,e2},
(hypothetical_on,S3,S1),
S3 ⊨ {n6,n7,n8,n9,n10,e3,e4,e5}}
Entities in the Domain: {mc,m,c,r,p1,t1,b,p2,t2}
{c2,m2,a,r2,p3,t3,p4,t4}

US4: [How shall [they] cross [the river]]?

S4 ⊨ { n11=(and,(3rd_person_pronominal,x),
(plural,x)),
n12=(and,(river,r3),(1,r3)),
e6=(and,(cross,p6,x,r3,t5),(future,t5)),
e7=(how,e6,Q?)}

Condition to be removed—> Result of removal:

(add,DisS3,S4) —>S4=S1. (The is difficult.) [Case5]

(3rd_person_pronominal,x)—>x=mc. [Case1]

(def,r3)	—>r3=r. [Case1]
(future,t5)	—>t5 is after the time of US4.

information state DisS4:

Facts:
{S1 ⊨ {n1,n2,n3,n4,e1,n5,e2,n11,n12,e6,
e7,Q?},
(hypothetical_on,S3,S1),
S3 ⊨ {n6,n7,n8,n9,n10,e3,e4,e5}}
entities in the Domain: {mc,m,c,r,b,p1,t1,p2,t2,p5,t5}
{c2,m2,a,r2,p3,t3,p4,t4}

Now the task for the puzzler is to substantiate the unknown fact Q? with a solution.

6. Discussion

It seems reasonable to assume that the last information state DisS4 contains all the necessary information for solving the problem—both explicitly as facts themselves and implicitly as something derivable from the facts and the common sense.

There is a lot of crucial information that only follows from the common sense, as McCarthy pointed out. For instance, u1 does not say explicitly that the group of missionaries and the group of cannibals "came together to the same side of the river", u2 does not say that the boat is available "on the bank of the river where they came", or that the boat seats two "normal size persons" and that all the six people are of normal size, and so on.

Even the intended meaning of the last question u4 requires a nontrivial set of commonsense assumptions. Its intended meaning is "How should the missionaries cross the river without being eaten by the cannibals?" In order to obtain this meaning from "How shall they cross the river?", the pronoun *they* must first be resolved. Although the missionaries would be the first candidate on the basis of grammatical saliency (i.e. mentioned with the previous matrix subject), the cannibals would almost equally be acceptable. However, the most natural referent is the group entity mc mentioned in u1—made accessible because u4 is a query about situation s1 rather than hypothetical situation s3.

The next hard task in understanding u4 is to infer the implicit connection among the capacity of the available boat, the hypothetical eating situation described by u3, and this question about the missionaries' crossing the river. One of the necessary assumptions to make here is that they will use the boat to cross the river, and by no other means such as swimming. A puzzle statement thus exemplifies a tightly circumscriptive reasoning in its extreme form. We assume that all and only what's relevant to the solution of the puzzle have been mentioned in the statement.

Even focusing on the information that language did make explicit, however, we realize how sloppy a job it is. This is so even for the most important information for the purpose of this text, namely, the number information.

Let us concentrate on u3: *If the cannibals ever outnumber the missionaries on either bank of the river, the missionaries will be eaten.* We know by definiteness resolution that *the cannibals* and *the missionaries* here correspond to the three cannibals and the three missionaries mentioned in u1. Since u3 evokes a hypothetical situation different from the situation described by u1, these definite referents do not exactly resolve to the groups of three members each. Rather, they can reference groups of any number between 0 and 3 that satisfy the IF-condition. Note that since these definites are within the scope of the indefinite *on either bank of the river*, the usual definiteness implication for existence is erased. Now, the assertion *x outnumbers y* is true if the number of x is more than the number of y. Then, in the case of the cannibals outnumbering the missionaries, we have the following six cases where it would be true:

	cannibals	missionaries
case 1	3	2
case 2	3	1
case 3	3	0
case 4	2	1
case 5	2	0
case 6	1	0

However, the missionaries cannot be eaten when their number is zero. This means that cases 3, 5, and 6 above are no danger, thus need not be avoided. All of these three cases in fact occur in Amarel's model solution above (the state 031 as case 3, 020 and 021 as case 5, and 010 as case 6). This means that a subset of the "outnumbering" possibilities must be taken advantage of rather than be avoided, which is not at all what the English sentence says. This shows that even in making such a crucial statement, language heavily relies on the world knowledge to get the intended message across.

Another observable sloppiness of linguistic description is in the use of grammatical plurality. Plural forms *cannibals* and *missionaries* are used as the cover term for all the above cases including those when the number is only one. This means that "PLURAL=2 or more" should not be taken too seriously in processing English.

These facts point to the essential need for a natural language processing system to be embedded in a commonsense reasoner. One such example is the abduction-based system of Hobbs et al. (1990), where a single inference mechanism of abduction applies on uniform logical forms representing both linguistic and non-linguistic information.

Now I will analyze the Japanese statement of the problem with the same approach.

7. Cross-linguistic Variation in What is Explicit

It is well-known that the Japanese grammar does not require the number and definiteness information of objects. Since number information is crucial in this puzzle, an interesting question is exactly how much of it is made explicit in the Japanese statement of the problem. I have obtained the total of ten "natural" translations of the English statement into Japanese. What I found was the following: (i) all the translations made the numbers of the missionaries, cannibals, and seating capacity of the boat explicit, (ii) only half of them made the number of the boat explicit, and (iii) none of them made the number of the river explicit.

This indicates that half of the people thought that common sense tells that there is only one boat available, and all of them thought that common sense tells that there is only one river involved. When the grammar does not require certain information to be explicit, we can directly see the amount of available commonsense information in this way.

Here is a typical translation into Japanese:

A Typical Translation Equivalent in Japanese

u1. Sannin no senkyousi to sannin no
3_person of missionary and 3_person of
hitokuizinsyu ga kawa ni yattekuru.
cannibal sbj river to come

u2. Hutari-nori no booto ga issou aru.
2_person_riding of boat sbj 1_vessel exist

u3. Mosi dotiraka no kawagisi de hitokuizinsyu no
if either of river_bank at cannibal of
kazu ga senkyousi no kazu
number sbj missionary of number
yori ooku naru to, senkyousi
more_than numerous become then, missionary
wa taberarete simau.
topic be_eaten will_have

u4. Dono you ni 0 kawa wo watareba yoi ka?
which manner in sbj river obj cross_if good Q

8. Conditions in Japanese

Although the general model of discouse should be the same even when different langugages are used, conditions generated from referring expression types and tense types are expected to differ to some extent. I find that the conditions for English referring expressions above can in fact be used to define the conditions for Japanese referring expressions relevant for the present analysis. Although this is by no means definitive, I find it highly promising.

Conditions in Japanese (Preliminary)

1. TOPIC (*NP-WA*)—same as English DEFINITE
2. BARE (e.g. *kawa*)

Case 1: same as English DEFINITE
Case 2: same as English INDEFINITE

3. ZERO PRONOMINAL

Case 1: same as English 3RD PERSON PRONOMINAL
Case 2: pre-condition: x is in the current US (indexical

Zero)
post-condition: none
4. NUM_no_N (e.g. *sannin no senkyousi*)
same as English INDEFINITE
5. N_no_NUM (e.g. *senkyousi no sannin*)
same as English DEFINITE
6. EITHER (*dotiraka no*)—same as English
7. PRES (Japanese PRES includes FUTURE)
Case 1: same as English PRES
Case 2: same as English FUTURE
8. PAST—same as English PAST
9. (ADD, i, s)—same as English

---

9. Analysis of the Japanese Text

What follows is a dynamic discourse analysis of the Japanese text above. Instead of Japanese predicates, I will use English predicates for an expository reason.

Note that the number information of object-type entities is treated uniformly as a property of a *set* entity. For instance, (missionary, x) specifies a set entity x with the "missionary" property, and x can be further restricted by quantity and unit properties such as (3, x) and (person_unit ,x) for *sannin* (three person_units).

initial information state DisS0:

Facts: { }
Entities in the Domain: { }

US1: [[[Sannin no senkyousi] to [sannin no hitokuizinsyu]] ga [kawa] ni yattekuru].

S1⊨{ n1=(join,{m,c},mc),
n2=(and,(missionary,m),(3,m),
(person_unit,m)),
n3=(and,(cannibal,c),(3,c),(person_unit,c)),
n4=(river,r),
e1=(and,(come,p1,mc,r,t1),(pres,t1))}

Condition to be removed—> Result of removal:

(add,DisS0,S1) —>S1 is in DisS1. [Case1]
(bare,m) —>m is in DisS1. [Case2-1]
(bare,c) —>c is in DisS1. [Case2-1]
(bare,r) —>r is in DisS1. [Case2-1]
(pres,t1) —>t1 coincides with the time of US1

information state DisS1:

Facts: {S1⊨{n1,n2,n3,n4,e1}}
Entities in the Domain: {mc,m,c,r,p1,t1}

US2: [[Hutari-nori no booto] ga issou aru].

S2⊨{ n5=(and,(rowboat,b),(1,b),(seat_of,b,2),
(boat_unit,b))
e2=(and,(exist,p2,b,t3),(pres,t2))}

Condition to be removed—> Result of removal:

(add,DisS1,S2) —>S1=S2. [Case2]
(bare,b) —>b is in DisS2. [Case2-1]
(pres,t2) —>t2 coincides with the time of US2.

information state DisS2:

Facts: {S1⊨{n1,n2,n3,n4,e1,n5,e2}}
Entities in the Domain: {mc,m,c,r,p1,t1,b,p2,t2}

US3: [Mosi [[dotiraka no kawagisi] de [hitokuizinsyu no kazu] ga [senkyousi no kazu] yori ooku naru] to, [senkyousi] wa taberarete simau].

S3⊨{ n6=(and,(number_of,k1,c2),(cannibal,c2)),
n7=(and,(number_of,k2,m2),
(missionary,m2)),
n8=(river_bank,a),
e3=(and,(more_than,p3,n1,n2,t3),
(location,p3,a),(pres,t3)),
n10=(missionary,m3),
e4=(and,(be_eaten,p4,m3,t4),(pres,t4)),
e5=(if,e3,e4)}

Condition to be removed—> Result of removal:

(add,DisS2,S3) —>(hypothetical_on,S3,S1) [Case3]
(bare,k1) —>k1 is in DisS3 [Case2-1]
(bare,k2) —>k2 is in DisS3 [Case2-1]
(bare,c2) —>c2 % c. [Case1-1] (c2 corresponds to but does not resolve to c.)
(bare,m2) —>m2 % m. [Case1-1]
(either,a) —>a normal river has exactly two banks
(topic,m3) —>m3=m2. (m3 resolves to m2.)
(pres,t3) —>t3 is sometime after the time of US3.
(pres,t4) —>t4 is sometime after t3.

information state DisS3:

Facts: {S1⊨{n1,n2,n3,n4,e1,n5,e2},
(hypothetical_on,S3,S1),
S3⊨{n6,n7,n8,n9,n10,e3,e4,e5}}
Entities in the Domain: {mc,m,c,r,p1,t1,b,p2,t2}
{k1,k2,c2,m2,a,p3,t3,p4,t4}

US4: [Dono you ni [0] [kawa] wo watareba yoi ka]?

S4⊨{ n11=(zero_pronominal,x),
n12=(river,r3),

Bare nominals are totally uncommitted to definiteness. I will assume the heuristics of preferring DEFINITE to INDEFINITE.

e6=(and,(cross,p6,x,r3,t5),(pres,t5)),
e7=(how,e6,Q?)}

Condition to be removed—> Result of removal:

(add,DisS3,S4) —> S4=S1. (This is difficult.) [Case5]
(zero_pronominal,x) —> x=mc. [Case1]
(bare,r3) —> r3=r [Case1]
(pres,t5) —> t5 is after the time of US4.

information state DisS4:

Facts: {S1 ⊨ {n1,n2,n3,n4,e1,n5,e2,n11,n12,e6,Q?},
(hypothetical_on,S3,S1),
S3 ⊨ {n6,n7,n8,n9,n10,e3,e4,e5}}
Entities in the Domain: {mc,m,c,r,p1,t1,b,p2,t2,p5,t5}
{k1,k2,c2,m2,a,p3,t3,p4,t4}

There are significant lexico-grammatical differences between English and Japanese, but both seem to function equally well as information tronsmission devices. This is no surprise if we assume, as in this paper, that the commonsense world knowledge underlies and supplements linguistic descriptions.

10. Conclusion

I have analyzed the missionaries and cannibals problem as a piece of discourse with a dynamic approach, and pointed out the areas where the linguistic contribution of information relies on the shared world knowledge. If we wish to see what the "discourse ontology" is, I would point at the information pieces and situations that define the shared information state at each point of discourse progression. More research is needed to make more precise claims about how the explict and implicit parts of the informationstate interact with each other and with updating operations triggered by each new utterance.

Acknowledgments

I would like to thank Johan van Benthem, Mark Gawron, Yasunari Harada, Jerry Hobbs, Yasuhiro Katagiri, Hiroshi Nakagawa, and Rebecca Passonneau for helpful comments on an earlier version of the paper.

References

1. ALLEN, J. Maintaining knowledge about temporal intervals, *Comm. ACM*, **26**(11) (1983), 832-843.
2. AMAREL, S. On representation of problems of reasoning about actions, In D. Michie (ed.), Edinburgh, Scotland: Edinburgh University Press. *Machine Intelligence*, **3** (1971), 131-171.
3. BARWISE, J. and PERRY, J. Situations and Attitudes, Cambridge, MA: MIT Press (1983).
4. van BENTHEM, J. Language in Action, **130**. Studies in Logic. Amsterdam: North-Holland (1991).
5. BRADY, M. and BERWICK, R. eds. Computational Models of Discourse, Cambridge, Mass. The MIT Press (1983).
6. DEVLIN, K. Logic and Information, Cambridge, England: Cambridge University Press (1991).
7. GÄRDENFORS, P. Knowledge in Flux: Modelling the Dynamics of Epistemic States, Cambridge, MA: MIT Press (1988).
8. GAWRON, M. and PETERS, S. Anaphora and Quantification in Situation Semantics, CSLI Lecture Notes Number 19, Stanford: CSLI Publications (1990).
9. GRICE, H. P. Meaning, *The Philosophical Review 66* (1957), 377-388.
10. GROENENDIJK, J. and STOKHOF, M. Dynamic predicate logic, *Linguistics and Philosophy 14* (1991), 39-100.
11. GROSZ, B. The representation and use of focus in a system for understanding dialogs, In the Proceedings of the Fifth International Joint Conference on Artificial Intelligence, Cambridge, Mass., 67-76, Los Altos: William Kaufmann, Inc. (1977).
12. GROSZ, B. and SIDNER, C. Attention, intentions, and the structure of discourse, *Computational Linguistics*, **12** (1986), 175-204.
13. HEIM, I. The Semantics of Definite and Indefinite Noun Phrases, Dissertation, Department of Linguistics, University of Massachusetts, Amherst.
14. HINTIKKA, J. and KULAS, J. The Game of Language, Dordrecht: Reidel (1983).
15. HOBBS, J., MARTIN, P., APPELT, D. and STICKEL, M. Interpretation as abduction, Technical Report. SRI International AI Center (1990).
16. HOBBS, J. Literature and Cognition, CSLI Lecture Note Number 21, Stanford: CSLI Publications (1990).
17. KAMEYAMA, M., OCHITANI, R. and PETERS, S. Resolving translation mismatches with information flow. In the Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics, Berkeley, California (1991).
Kameyama, M., Passonneau, R. and Poesio, M. Temporal Centering (in Preparation).
18. KAMP, H. A theory of truth and semantic representation, In J. Groenendijk, T. Janssen, and M. Stokhof, eds., Formal Methods in the Study of Language, Amsterdam: Mathematical Center (1981), 277-322.
19. LEWIS, D. Scorekeeping in a language game, *Journal of Philosophical Logic*, **8** (1977), 339-359.
20. MCCARTHY, J. Circumscription—A form of non-monotonic reasoning. *Artificial Intelligence*, **13**(1, 2) (1980), 27-39.
21. PARTEE, B. Nominal and temporal anaphora, *Linguistics and Philosophy 7* (1984).
22. PEREIRA, F. and POLLACK, M. Incremental interpretation, *Artificial Intelligence* (forthcoming).
23. POLANYI, L. A formal model of discourse structure, *Journal of Pragmatics*, **12** (1988), 601-638.
24. REICHENBACH, H. Elements of Symbolic Logic. New York: Macmillan (1947).
25. SCHA, R. and POLANYI, L. An augmented context free grammar of discourse. In the Proceedings of the 12th International Conference on Computational Linguistics, Budapest, Hungary (1988), 22-27.
26. SIDNER, C. Focusing in the comprehension of definite anaphora, In M. Brady and R. Berwick, eds. (1983), 267-330.
27. THOMASON, R. Accomodation, meaning, and implicature. In P. Cohen, J. Morgan, and M. Pollack, eds., *Intentions in Communication*, Cambridge, Mass.: The MIT Press (1990), 325-364.
28. WEBBER, B. So what can we talk about now? In M. Brady and R. Berwick, eds. (1983), 331-371.

(Received December 2, 1991)