Toward a Rational Model of Discourse Comprehension

Jerry L. Morgan
Center for the Study of Reading
University of Illinois
at Urbana-Champaign

I.   Introduction

I begin my tale with the moral: a quotation
from the greatest English grammarian, Otto
Jespersen (1965).

> The essence of language is human activity--
> activity on the part of one individual to
> make himself understood by another, and
> activity on the part of that other to
> understand what was in the mind of the first.
> These two individuals, the producer and the
> recipient of language, or as we may more con-
> veniently call them, the speaker and the hear-
> er, and their relations to one another, should
> never be lost sight of if we want to under-
> stand the nature of language and of that part
> of language which is dealt with in grammar.
> But in former times this was often overlooked,
> and words and forms were often treated as if
> they were things or natural objects with an
> existence of their own--a conception which
> may have been to a great extent fostered
> through a too exclusive preoccupation with
> written or printed words, but which is funda-
> mentally false, as will easily be seen with
> a little reflexion. (p. 17)

But the temptation to think of language as pure
form is great; Jespersen himself slips into this
metaphor a few pages later:

> . . . we always find that there is one word
> of supreme importance to which the others are
> joined as subordinates. This chief word is
> defined (qualified, modified) by another word,
> which in its turn may be defined (qualified,
> modified) by a third word, etc. (p. 96)

But words do not define, modify, or qualify other
words. Speakers define, qualify, and modify.
This confusion is so tempting that it is perva-
sive in every field that studies language, at any
level. It is almost universal in linguistics. We
find it, for example, in the following from
Halliday and Hasan (1976), who probably know
better:

> Let us start with a simple and trivial exam-
> ple. Suppose we find the following instruc-
> tions in the cookery book:
>
> [1:1] Wash and core six cooking apples.
> Put them into a fireproof dish.

It is clear that them in the second sen-
tence refers back to (is ANAPHORIC to) the
six cooking apples in the first sentence.
This ANAPHORIC function of them gives cohe-
sion to the two sentences, so that we inter-
pret them as a whole; the two sentences
together constitute a text. Or rather, they
form part of the same text; there may be more
of it to follow.

> The texture is provided by the cohesive RELA-
> TION that exists between them and six cooking
> apples. It is important to make this point,
> because we shall be constantly focusing at-
> tention on the items, such as them, which
> typically refer back to something that has
> gone before; but the cohesion is effected not
> by the presence of the referring item alone
> but by the presence of both the referring
> item and the item it refers to (p. 2).

There are two serious confusions here. First,
words do not refer; speakers refer to things by
using words. The word them does not refer to
anything at all, obviously so since it can be used
to refer to any set one wants to refer to. There
is no particular set of entities that one can say
the word them refers to. But one can use it to
refer to sets of things, when one's intended
referent will be recoverable in some way by the
hearer.

The second confusion is the idea that words
"refer back" to other words. The muddle here is
obvious. Whether it is people or words that
refer, it is things, not (usually) other words,
that they refer to. Thus in Halliday and Hasan's
example [1:1], it is not the words six cooking
apples that them is used to refer to; one is not
being instructed to put three words in a fireproof
dish. The word them is used to refer to certain
apples that were previously referred to by use of
the words six cooking apples. My objection to
such descriptions is not based merely on a nig-
gling concern with sloppy language. If it were,
one might respond that it's clear what Halliday
and Hasan mean here, so my complaint is beside
the point. Rather, I think the pervasive confu-
sion on just this point is a symptom of a serious
comceptual confusion that renders a lot of the
related work useless. This is the case with the
passage from Halliday and Hasan. They say that it
is some relation between sentences in a text that
gives it "cohesion", that renders it coherent,

"so that we interpret them as a whole; the two sentences together constitute a text." The relation that gives rise to this cohesion is that _them_ in one sentence "refers back" to the _six cooking apples_ in a previous sentence. If _we_ interpret this phrase charitably, then the question arises, how do we know what _them_ refers to? How do we know that it refers to the apples, and not to two of the writer's bachelor uncles? We can't _know_ such a thing. We can only assume that the writer is rational, and that the recipe is coherent. If it is coherent, we are justified in assuming that it is the apples that are referred to by _them_. But there is a vicious circularity here. The recipe has cohesion, is a coherent text, just in case _them_ refers to the apples. But we are only justified in inferring that _them_ refers to the apples if we _assume_ that the text is coherent. Thus, in spite of Halliday and Hasan's claim, it is not the anaphoric facts that give rise to cohesion; rather, the assumption that the text is coherent gives rise to the inference that _them_ refers to the apples.

This kind of confusion, it seems to me, arises from the linguist's habit of looking at every aspect of language in terms of linguistic forms and relations between them. Thus in this case the mistaken characterization of reference as a relation between words, and of coherence as a property of an abstract _linguistic_ object called a text. In the rest of this brief paper I want to sketch an opposing view, and to claim that notions like "reference," "text structure," "relevance" and "coherence" are best treated, at least in part, in terms of _communicative acts_ and the plans and goals of speakers/writers who perform such acts.

II. Three Ways of Looking at a Text

Assume for the moment that we know what a text (oral or written) is, and can tell a coherent text from a random transcription of English sentences (I will return to what counts as a coherent text later). Then there are (at least) three kinds of things and relations involved in a text.

1. _Sentences_. First, conventional wisdom in linguistics has it that texts consist of sentences. I shall accept this for the moment, though a bit later I will show cause to modify it.

But what kind of "thing" is a sentence? It is, if anything is, an abstract linguistic object, a unit of form. It is not a proposition, nor a fact, though it is a means by which such things are asserted, denied, questioned, etc. Nor is a sentence a speech act, though a speech act will usually be performed by means of the utterance of a sentence. But a sentence and an utterance of a sentence are different kinds of things.

A sentence is not the kind of thing that is true or false; "facts," or "propositions," that sentences can be used to express, are true or false. Or perhaps it would be more appropriate to speak of _assertions_ as being true or false. At any rate, it is quite clear that it is nonsense to speak of sentences as true or false, as

evidenced by the familiar problem of indexical expressions.

A sentence, then, may be used to assert that something is true, or false, or has occurred, but the sentence itself is not true or false, and does not occur. Thus relations like causation, order in time, entailment, and so on, do not hold between sentences. It is not clear what kind of relation can accurately be said to hold between the sentences of a text.

2. "_Facts_." The second kind of "thing" involved in a text is what I shall call "facts." (Notice that I do not say texts _consist_ _of_ or _contain_ facts; merely that they somehow involve facts.) The term "fact" is a bit misleading--though I can think of no better term--in that I wish to include as facts events, states, and so forth that do not actually hold in the real world; "propositions," more or less.

Relations among the "facts" involved in a text, then, consist of two classes: first, the same relations that hold between facts in the real world--causation, relations of temporal order, motivation, and so forth; second, those relations that have to do with logic and hypothetical facts, like entailment and contradiction. It may be necessary to distinguish facts on the one hand and propositions on the other, on grounds that relations between facts are of a kind different from relations between propositions, but I will ignore the problem here.

3. _Speech acts_. The third kind of thing involved in texts is the "speech act" (by this term I mean to include as a sub-case acts of linguistic communication by writing). Speech acts are not sentences, nor "logical forms," nor propositions, in spite of occasional attempts to define them in these terms. They are acts, just as the term implies.

Relations between the speech acts involved in a text are just those that can hold between acts in general. First, since an act is a sub-type of event, the relations that can hold between events can, in general, hold between acts, thus between speech acts: relations of temporal order, for example. Second, since a speech act is a sub-type of act, relations that can hold between acts can, in general, hold between speech acts. The most important relation in this regard is the relation of purpose: one does such-and-such _in order_ that such-and-such; or one performs a certain act in order _thereby_ to perform a second act. Long chains of these relations can hold between acts. I may throw a switch in order to turn on a light in order to frighten away a burglar in order to save the family jewels. I may tell my friend that there is a charging bull behind him in order that he realize that he is in danger, in order that he get out of the way. It is a mistake to ask whether my speech act was an assertion or a warning, since this presupposes that the two are mutually exclusive. It was both; I asserted something and _thereby_ warned somebody, just as I threw the switch and _thereby_ turned on the lights. I may make a certain mark on a piece of paper, thereby marking my ballot for Smith, thereby casting a vote for Smith. I may

assert that I will do the dishes, thereby volunteering to do the dishes. And so on.

It is commonly the case that acts are linked by complex relations of purpose and goal, including the case where one act is performed by means of performing another act. This is especially true of communicative acts.

There are several subvarieties of speech acts, for which several taxonomies have been proposed; Austin (1962), McCawley (1977), for example. One important distinction in kind is the distinction between the act of saying a sentence, and the act one thereby performs. In performing an act of saying the English sentence "Your hair is in my yogurt" I may, in the right circumstances, thereby inform someone that their hair is in my yogurt. The first kind of act, the act of saying, includes making sounds in a way that conforms to the conventions for what counts as a saying of a sentence, or making visible marks in a way that counts as a saying of a sentence. Texts, then, do not really consist of sentences, but of sayings ("uses") or sentences; or in the case of written texts, of a permanent kind of record of uses of sentences.

III. The Interpretation of Texts

A. Speech acts. The interpretation of a text, then, consists of the interpretation of this record of sayings of sentences. Each saying is interpreted in terms of some speech act(s) performed by saying a given sentence. (Henceforth by "speech act" I will mean the communicative act one performs by saying a sentence, as opposed to the act of saying itself.) There are three aspects to the interpretation of speech acts: the interpretation of what speech acts are performed-- assertion, promising, denial, questioning, warning, etc.--by the saying of the sentence; the interpretation of what "facts" are asserted, denied, etc.; and the interpretation of the speaker's purpose and goal in performing the speech act.

As an aside I should mention the special instance where nothing is directly asserted, denied, etc.: the case of speech acts of reference. An act of asserting, etc. (for brevity I will henceforth use assertion as representative of all speech acts types), will usually include an act of referring as a subpart; a reference to the entity of which something is asserted. But acts of referring can occur independently. For example, I might say "The door!" to someone under a number of circumstances, to get them to open it, close it, shoot the bad guy standing in it, or merely observe what beautiful hardwood it is made of. It would be a mistake to say that "The door!" means any of these things, or that I have performed (directly) any kind of speech act beyond merely referring. I have only referred to the door, thereby to call my hearer's attention to it, with the expectation that when he turns his attention to the door he will realize what it is I want him to do about it.

The typical immediate goal associated with speech acts of all kinds is the same: that the hearer modify his model of a certain "world"

(in the sense of "possible worlds") in a way that involves the "facts" that are asserted, etc. in the speech act. The world involved may be the real world, or, in the case of story-telling, for example, some imaginary world. The modification may include the construction ex nihilo of some hypothetical or imaginary world. The relation between the "facts" of the speech act and the intended modification vary with the nature of the speech act; but in all cases some modification is involved. The simplest case is that of assertion; normally the immediate goal of an assertion is that the hearer modify the world under discussion in a fashion that makes the asserted fact true in that world. In the case of yes-no questions, the goal is that the hearer modify his model of the world such that in that world the speaker wants the hearer to tell him whether the fact questioned is true. In the case of imperatives, the goal is that the hearer modify his model such that in that world the speaker wants the hearer to bring about the truth of the ordered fact, and that certain social consequences will follow from non-compliance.

The raw datum of comprehension, then, is not the sentence or the proposition, but the fact that a certain speech act has occurred. In comprehension, people do not process sentences as abstract formulae; they observe that someone has said something to them, and attempt to interpret that act and its consequences, which may include modification of their model of the world. The process of modifying the model according to what is said is not direct, but the result of several steps of evaluation. Interpretation of an assertion might go roughly like this, from the viewpoint of the hearer (where S is the speaker, A the addressee; addressee and hearer may be identical):

> S has said x to A. Saying x counts as asserting p. S knows that saying x counts as asserting p. S knows that therefore his saying x is likely to be interpreted by A as an assertion of p. S has done nothing to prevent A from making this conclusion. Therefore S's intention is that his saying x be taken by A as an assertion of p. Then if S is sincere S believes that P is true. A must conclude that S has asserted p because he wants A to take p as true and modify his model of the world accordingly.

But the decision to believe p, i.e. modify his model of the world to include p, is a matter of choice on H's part, not an automatic consequence of processing the "sentence." The steps involved in making this decision are equally complex, involving the ability to construct a hypothetical world just like the real one except that p is true, to evaluate the consistency and plausibility of that world, and so on. Some of the facts that are asserted will relate to this decision-making process. For example, in saying (1) my goal is most likely that the hearer come to believe that both facts asserted are true.

(1) John is here. He has a dog with him.

But in the case of (2), I am not so much con-
cerned with the second asserted fact in itself,
but with the goal that from concluding that it is
true, the hearer will be more likely to believe
the first, since I intend that he take the second
fact as evidence that my source is reliable.

(2) The world is flat. It says so
in the Encyclopedia.

Matters that are sometimes construed as rhetorical
relations between sentences fall into this cate-
gory. Some fact is asserted not because it is
important in itself, but because it bears on H's
evaluation of some other asserted fact. Thus the
relation is not one between sentences, but be-
tween speech acts. One speech act is performed
in order to influence the interpretation and
evaluation of another. At any rate, my point here
is that in comprehending a text in the serious
sense, comprehension proceeds not from some dis-
embodied abstract object called a "sentence," nor
from a "proposition," but from the perceived fact
that S has said such-and-such, and that so saying
counts as a speech act of a certain type.

There is another way in which modification
of the world model is not a direct function of
the asserted fact: the widely studied problem
of inference. Given the hearer's acceptance of
what the speaker has asserted, incorporation of
the facts into the model of the world may involve
more than merely adding the asserted facts.
There is, for example, a general principle of
ceteris paribus that comes into play in considera-
tion of alternative worlds. Roughly, when con-
structing a model of a world alternative to some
point-of-reference world (usually the real one),
the hearer will assume, lacking evidence (from
assertion or inference) to the contrary, that the
alternative world is consistent with the point-of-
reference world in all relevant respects. To take
an extreme example, if someone is telling me about
life on Arcturus, I will assume that the laws of
physics are the same there as on earth, unless
something the speaker says leads me to believe
otherwise. In the same way, hearers will assume,
lacking counter-evidence, that what is typical
in the point-of-reference (e.g. real) world is
also typical in the alternative world. They will
also assume that things of a given type have the
properties typical of things of that type.
Gricean rules of conversation support these in-
ferential strategies in the following way: The
hearer knows that the speaker knows the hearer is
likely to make inferences according to these and
other strategies. The speaker has done nothing
to prevent the hearer from making them. There-
fore the hearer is justified in inferring that
the speaker intends for the inference to be made.

Using these and other strategies, then, the
hearer modifies his model of one or more worlds,
based not on detached sentences or propositions
floating in some abstract semantic space, but on
his observation that a certain person has per-
formed a certain speech act.

B. Relations among speech acts. But there
is more to the interpretation of a text than just
the interpretation of individual speech acts. A
speech act is performed for some purpose, with
some goal in mind. And complete understanding
of a text involves the ability to infer such
goals and purposes at every level, from in-
ferring the purpose of referring expressions to
inferring the speaker's overall goal in con-
structing the text. One can understand every
sentence in a text, yet come away puzzled at what
it was the speaker was trying to say, or what
the parts of the text had to do with each other.
To understand the purpose of a speech act is to
understand how it relates to a goal, how it is a
step toward the achievement of that goal. The
most appropriate kind of theory for this aspect
of a text is a theory of plans, in which pur-
poses, goals, acts, and intentions play a crucial
role.

There are a large number of goals a speaker
can have in constructing a text, including many
that are irrelevant to comprehension: to derive
royalties, for example, or to confuse an enemy
by furnishing misinformation. A proper theory
of text comprehension must distinguish goals
like these from those that are central to com-
munication and comprehension, probably by means
of conditions like those Grice (1957) proposes
as criteria for meaning.

C. What can go wrong. Then we can sketch
the task of text comprehension as follows:

1. From the sounds or markings, H must
   recover what sayings are recorded in
   the text, in what order.

2. From this H must recover what speech
   acts have been performed, in what
   order.

3. From each speech act H must recover
   what facts are being asserted, denied,
   promised, etc.

4. From this H must infer what modifica-
   tions he is intended to make in his
   model of the world, and how to make
   them in the most consistent way;
   this is not a direct function of
   the facts, as discussed earlier.

5. For each speech act H must infer a
   purpose that is consistent with the
   purposes he inferred for earlier
   speech acts; or he must revise
   earlier hypotheses about purposes
   accordingly. Questions H must infer
   answers to are, "Why did the speaker
   perform this particular speech act,
   at this particular point in the text?"
   and "Why does he want me to have
   this particular fact just now?"

6. From speech acts and their purposes
   taken jointly, he must construct a
   hypothesis of the speaker's goal in
   the text, and of the plan that the
   speaker is following in advancing
   toward that goal. At each step the
   purpose of a given speech act must
   somehow be construed as consistent

with, and actually advancing that plan, or the plan hypothesis must be modified so that it can.

7. From hypotheses about the speaker's plans and goals in the text, the hearer will form expectations: hypotheses about what the speaker is likely to do next in advancing toward the goal of the text.

These matters do not proceed in separate compartments, of course, but feed each other. The plan one has constructed so far can influence decisions about what speech act is performed in a given utterance, for example, and the interpretation of pronouns can be influenced by hypotheses about the speaker's goals, just as a decision about what a referring expression is being used to refer to can influence the process of inferring a plan, and expectations about what the speaker will do next can influence the interpretation of what he actually does.

From this sketch we can derive a picture of where things can go wrong in comprehension, giving some insight perhaps into notions like "text structure," "relevance," and "coherence."

The hearer can have difficulty in tasks 1 through 3, of course, but the matter seems straightforward, so I will not discuss it. Difficulties can arise in task 4 in at least two ways. First, the world described may be so factually or logically bizarre, or so inconsistent with the hearer's beliefs (a description of ping pong in a black hole, for example), that the hearer is unable to construct a consistent model with any degree of detail. The term "incoherent" might be applied to such cases, but I think this is not what linguists mean by "textual coherence," which I will discuss below.

A second kind of difficulty with task 4 arises when the facts are consistent, but the hearer lacks the knowledge necessary to figure out how to construct a consistent model that incorporates those facts. For example, if I describe in detail a walk through the South Side of Chicago, a person who has been there before will be able to construct a much more richly detailed model of my walk than a person who has not.

Difficulties can arise with task 5, insofar as the hearer is able to understand clearly what's being asserted, but unable to determine the speaker's purpose in asserting it. Here is the place to look for an adequate definition of relevance. Actually there are two senses of the word in ordinary usage. One can speak of relevance as a relation between facts. One fact is relevant to another when the truth of one depends in some way on the truth of the other. But I think more often, linguists who speak of "relevance" as a problem of text comprehension have in mind a problem that is best treated in terms of purposes behind speech acts. Given a hypothesis about the goal and plans of a speaker in a text, a given "sentence" (i.e. speech act) is taken to be irrelevant when the hearer is unable to see how it functions within the plan to

advance toward the goal. Relevance under this interpretation, then, is a relation between an act and a goal, not a relation between sentences. If in recounting my recipe for Wienerschnitzel I describe my new driveway, it's not that the sentences are irrelevant; rather, I have done something irrelevant. The same passage may count as full of irrelevancies, relative to one goal, but uniformly relevant, relative to another goal.

Task 6 is probably the most complex and difficult, and the one we know least about. But I suspect that it is a likely source of progress in understanding such important but elusive notions as "coherence," "text structure," and "topic." In understanding a text, the hearer unconsciously searches out a primary goal behind the text, and tries to construe every part of the text as a purposeful step toward that goal, according to some plan. If the hearer is unable to reconstruct the goal or plan, or indeed decides there is none, the text will be judged "incoherent." Coherence is not a formal property of texts, nor of "logical structures" of texts, but a function of the hearer's ability to relate parts of the text to a plan for achieving some goal. If it should turn out that the coherence of texts correlates with the number of pronouns, it would be a mistake to conclude that lots of pronouns makes a text coherent. Rather, it would show that coherent texts tend to be ones where the speaker says a lot about one or two topics, rather than saying one thing about 32 topics. It is the coherence of what the speaker is doing in the text that gives rise to the abundance of pronouns; the formal property of having a lot of pronouns does not give rise to coherence.

At least some aspects of "text structure" can also be treated in these terms. An ideal unified paragraph, for example, is a unit of function, not of form; the speaker formulates a subgoal as a step toward the primary goal of the text, and sets about to achieve that goal in a series of speech acts. Insofar as the hearer is able to discover this, the series of speech acts will be judged to be a unit; but a unit of function, not of form, defined not in terms of sentences or propositions, but communicative acts of some person, who uses those sentences to convey those propositions.

It is likely that an understanding of task 6 will lead to an understanding of "topic" as well. At present, there are nearly as many definitions of "topic" as there are linguists, and none of the definitions is clear enough to be usable. For some linguists the topic is a certain NP in a sentence; for others a topic is something a sentence has, though the NP may not be present in the sentences. For some every sentence has a topic; for others, only some sentences have topics. But I suspect that all of these attempts miss by a wide mark. First, it is not NP's that are topics, but the things in the world they refer to. Second, I suspect that such definitions can never be made sense of in that it is speakers, not sentences or even texts, that have topics. If so, then the proper theoretical treatment of "topic" would be framed

in terms of a theory of complex communicative acts, not formal linguistic properties.

IV. Conclusion

In this speculative paper I have proposed a way of looking at the comprehension of connected text that is counter to the linguist's usual way of looking at language. My main point is that certain notions are more likely to receive adequate treatment in a theory that incorporates a theory of speech acts, a theory of plans and goals, and a theory of inference, in place of a theory that looks for answers in terms of formal properties of texts. It remains, of course, to develop such theories to a level where my claims can be rigorously tested. The construction of such theories should be a prime goal of theoretical linguistics.

References

Austin, J. How to do things with words. Oxford: Oxford University Press, 1962.

Grice, H. P. Meaning. Philosophical Review, 1957, 66, 377-388.

Halliday, M. A. K., & Hasan, R. Cohesion in English. London: Longman, 1976.

Jespersen, O. The philosophy of grammar. New York: Norton, 1965.

McCawley, J. Remarks on the lexicography of performative verbs. In A. Rogers, R. Wall, and J. Murphy (Eds.), Proceedings of the Austin Conference on Performatives, Presuppositions, and Implicatures. Arlington, Va.: Center for Applied Linguistics, 1977.

Footnote