4-2-1993

A Semantic Theory of Information Recovery

Paul Baltes
A Semantic Theory of Information Recovery

Paul Baltes

Ellipsis is the ancient crux of the theories of language (translation mine).
Buhler as quoted in Hernandez (1984)

Introduction and Foundation

The purpose of this paper will be to investigate some of the ways in which people decode and retrieve information presented to them in discursive texts, both spoken and written. Specifically the paper will examine information which is ellipsed or deleted (and claimed to be fully recoverable) and the processes in which speakers and hearers engage to present and recover such information.

Communication is a series of structured discursive practices. When we know the language of the communication, we know the rules underlying these potential and possible structures and are able to interpret them to understand meaning. Communication, however, is not a simple aggregation of elements to convey meaning. Much of what is communicated is done so through context, through linguistic convention or knowledge, or through extralinguistic information we are able to extrapolate onto the text. The interpretive recovery which takes place as hearers attempt to understand the communicative intents of the speakers is an integral part of communication.

Defining the Problem

The term ellipsis is often constrained in linguistics to discuss information which is deleted for reasons of economy, emphasis or style, but which is completely and unambiguously recoverable from linguistic context (cf. Crystal 1991; Quirk et al. 1985). In this treatment, however, I will be broadening the term to include cases of omissions or deletions not handled within such constraints including but not limited to reduced, contracted, abbreviated, or deleted constructions.

In choosing what information to include in the discourse, speakers often make decisions (although not completely on a conscious level, c.f. Dosher and Rosedale [1991]; Gerken [1991]; Bloom [1990]) concerning which utterances and clues to provide based on what they believe the hearers already know and/or can easily reconstruct. To do anything differently would violate accepted maxims of conversation and cooperation. Information which is intended and yet not expressly provided is ellipsed or reduced by the speaker. Even the most cursory glance at transcripts of conversations reveals a frequent use of ellipsis (Ricento 1987) by native speakers. Printed texts, meanwhile, which may often focus on instructing or conveying information pose problems for speakers and especially for non-native speakers in that a great deal of content is ellipsed on the assumption it is fully recoverable.

Philosophers of language (Austin, Grice and Searle for example) have long argued that the speaker's intentions are inextricably woven into the fabric that is communication. Attardo (1992) presents a somewhat simplified view of this process: a speaker (S) has a certain meaning (M) in mind that he/she wishes another speaker (the hearer, H) to have. By uttering a sequence of sounds, which are connected to certain meanings (utterances), S attempts to communicate to H his/her intended M, by providing enough clues for H to "reconstruct inferentially M." This process is then repeated over and over as each S becomes an H when the first H responds to the original M from the first S. A similar process occurs for the dialogue between written texts, their authors and their readers. This process differs, however, in the amount of time readers may have to process the text and in the differing linguistic system involved for writing. Nevertheless I will use the term speaker to refer to both conversational speaker and author, and the term hearer to refer to conversational hearer and audience of a written text.

The only constraint which all theories which claim to be able to effectively treat ellipsis and re-
duced information place on exactly what information can be ellipsed is that such information must be fully recoverable by the hearer. Not only is this a dramatic aspect of syntactic theories, but it is also an important element underlying Grice’s conversational maxims. Speakers and Hearers engaged in communication assume in interpreting discourse that the communicative structures are relevant, true, and contain as much information as is necessary for them to be understood.

Thus far little has been done in an attempt to understand what goes on in the hearer’s recovery processes to interpret the ellipsed or reduced information. Linguistics and philosophy of language acknowledge that this operation does occur but focus on other aspects of the communicative process. Levinson (1985) notes that such a theory of ellipsis and recovery is necessary to understand the workings of communication, but that there has been insufficient progress in formulating one. Kinneavy (1971) reveals that in understanding informative discourse, only tentative and sketchy attempts have been made to illustrate the semantic processes involved, and these have been made on artificial miniature systems with extrapolations to the larger communicative process. Kato (1986), meanwhile, declares that studies which examine the decoding of the material which will prove no trouble to the native speaker, but which may pose some minor problems for the non-native speakers (although any such difficulty would most likely not affect the understanding of the content of the discourse).

While these examples do not present the full scope of syntax in dealing with ellipsis, they are representative of the types of data and examples found in the above sources (cited in 1.2). These examples, however, represent the lower bounds of the scope of this investigation. Ideally any theory of recovery would have to take into account every deletion from the phonemic level to that of metaphor and symbol.

Syntactic Treatment of Ellipsed Information

While there has been a significant body of research done in the name of ellipsis or recovery, most of it describes a specific syntactic notion dealing with a part of the structure which has been eliminated (cf. Chao [1988]; Lobeck [1991]; Clifton et al. [1991]; Piccoli [1988]; Hernandez [1984]; Kuno [1982]; Sag [1977]) and how a particular syntactic theory recovers the ellipsed information. Examples of such treatments include (1-5):

(1) Fred went to the mall, and Robin //1 to the rally.
(2) Fred went to the mall, and so did Robert //.
(3) Fred went to the mall and Robin did // too.
(4) Fred went to the mall, and Robin didn’t //.
(5) Did you go to the mall?
   Yes, I went //.
   Yes, I did //.
(6) Is he really going to buy a car. He owns three // now.

These examples represent several types of structural or strict ellipsis. In example (1), only a single word has been ellipsed. In (5) an entire phrasal constituent has been ellipsed. Such examples are recoverable if the hearer understands the grammatical relationship involved which enables the repeated structure to be ellipsed and fully recovered. In (2), (3), (4), and (6), there are some slight grammatical changes (tense, plural) on the recoverable material which will prove no trouble to the native speaker, but which may pose some minor problems for the non-native speakers (although any such difficulty would most likely not affect the understanding of the content of the discourse).

In examining the small but significant body of research dealing with syntactic treatments of ellipsis, I shall discuss its contribution to our understanding of ellipsis and recovery, as well as describe where semantics has been smuggled into the research and yet never identified as being anything other than syntax. The dissertation will primarily be concerned with common examples of discursive ellipsis such as those found in everyday texts as such practices constitute the greater frequency of occurrence and also have received the least attention.

Semantic Basis for a Treatment of Ellipsis

There is a great deal more information which must be recovered by audiences which syntactic treatments alone cannot account for, as in (6-9):

(6) Chelsea was a dime short and had to do without milk.
(7) Chelsea was a dime short and had to do without family.
(8) She saw a black cat and turned around and went home.

1 The double slashes indicate the place from which information is ellipsed.
These examples (6-8) from Raskin [1985]) are fully realized in terms of their grammatical components and yet there is additional information required for the hearer to understand what the utterances mean. Raskin (1985) illustrates his script-based semantic theory where the native speaker is able to construct contexts of utterances (aside from those readily available from the situation) from encyclopedic information and world knowledge. Such scripts enable us to understand the money-commodity relationship in (6) and the black cat-bad luck connotation for (8). Other information may also be reduced or hidden in metaphor (cf. Lakoff and Johnson [1980], and Tourangeau and Rips [1991]), symbolism, or even jokes (Raskin [1985]). Certainly syntax alone cannot help us understand these examples or even why a particular utterance would be funny or not (see Attardo et al. [to appear]).

The interpretive reconstruction which takes place as hearers attempt to understand the communicative intents of the speakers is thus an integral part of communication.

What will be recoverable, however, will vary because of culture and linguistic proficiency. Most speakers assume native proficiency in the production of texts, and even when the text is directed towards a non-native speaker problems may arise due to the linguistic conventions or strategies inherent in the discursive practices of the native speaker.

Much of the work undertaken in formulating theories of conversation and communication is based on Grice's notion of a co-operative principle (briefly discussed above). This principle proposes ways in which speakers co-operate to facilitate communication. Grice sought to identify the principle which people actually use in conversational exchanges in his conversational maxims. The Maxims of Quantity (make the contribution as informative as required for the current purposes of the exchange; do not make the contribution more informative than is necessary) and Relevance (make the contributions relevant especially help us to understand how people unconsciously process remarks and offer the only way in which we can understand utterances such as (10-11):

(10) A: Where's Kevin?
   B: There's a yellow VW parked outside Sheila's house.
(11) A: Want to go to the movies this afternoon?
   B: There's a sale at Penny's.

According to the principle of cooperation, A assumes that B's remarks in each case are relevant and that B has provided just enough information to answer the questions asked. We can understand that Kevin is known to drive a yellow VW and therefore may be in Sheila's house and that speaker B does not want to go to the movies because of a sale at JC Penny's.

Of course people are constantly involved in the trying to decide what information is relevant and how much to provide those with whom they converse. There must be a continuum from "old," "given" or less important information (theme) and "new" or more important information (rheme) (cf. Hernandez [1984]; Needham [1990]; Attardo et al. [to appear]). Flower, for example, (1979; also see Bloom 1990) illustrates that the egocentric talk of the child, or the inner voice which we use as adults (as is writing in a diary or a journal) is highly elliptical in that we eliminate reiterating information we already know (thematic) and focus on only new predicates (rheme). Explicit referents and subjects are usually absent, for example. This occurs despite the high relevance of the information to the rhetor and the situation of the text. Such clues need to be further investigated as to how speakers decide what to ellipsis based on what they assume audience members already know.

Halliday and Hasan (1976) and subsequent work by Witte and Faigley (1981) focus on the notion of cohesion or the semantic relationships within and between/among sentences which establish their relation to each other (cf. Raskin's notion of semantic recursion [1985], and work done by Conine et al. [1991]). The authors primarily focus on three types of grammatical cohesive ties—reference, substitution and ellipsis. At times these cohesion relationships add new information to the sentences, (with comparatives, for example). Similarly, what is ellipsed often influences the interpretation of the discourse.

This work as well as other work in coherence provides the most substantial foundation for the work I'm pursuing in building a theory of recovery. I see a great overlap in the description of processes which speakers and writers engage in to interpret, understand and even recover ellipsed or deleted information. Such processes include cohesion, coherence, redundancy, semantic recursion, subcategorization, entailment, scripts, selectional restriction rules, presupposition, implicature, anaphora, cataphora, theme/rheme distinctions, local reduction, local synonymy (local in these two instances refers to the ungeneralizable reduction or establishment of a synonym for the immediate text only), paraphrase relationships, transformations, ability to recognize and manipulate phrasal constraints (ala Chomsky, 1957, Syntactic Structures), locus of Moses errors (Reder & Kusbit 1991), semantic distance, relevance judgments, allusions, mode of discourse, speech genre, governance, the ECP (empty category principle), context, ambiguity resolution, inferences, connotations, metaphors and symbolism interpretation, extralinguistic information and world knowledge. I will show that much of the workings of these "processes" are similar in function and provide the basis for my exploration.
of the semantic triggers involved in recovering ellipsed information.

Rules of Semantic Recovery of Ellipsis

We will begin to discuss the strategies which native and non-native speakers use to recover ellipsed information in terms of the following example:

(12) A: “The Korean jet shot down by the Soviets was a Spy plane.”
B: “With 269 people on board?” (Carberry 1989)

If ellipsed information can be reconstructed from the cues in the (surface) structure of the text (as most syntactic theories suppose), then the reconstruction of (12) would look like (13):

(13) B: “Was the Korean jet shot down by the Soviets a spy plane with 269 people on board?”

Instead the more appropriate recovery (which would include the original intention of the second speaker (B)) would look more like:

(14) B: “How can you think that the Korean jet shot down by the Soviets was a spy plane, when it had 269 people on board?”

Most speakers have little or no difficulty recovering the intention of speaker B, but this is a very demanding example for others, including the non-native speaker.

Here we see just one example of where syntax fails in most all cases but those of strict elliptical demanding example for others, including the most speakers. We see a few examples will help to clarify some of these strategies.

(15) John's been drinking again. John don't take a drink tonight.
(16) The eagle is the lion among birds.
(17) Men are wolves.

(18) Q: What were Ted and Lisa doing in the library?
A: Well, they weren't studying.
(19) When Oral Roberts told his followers that if he didn't receive $8 million for his ministry, God would “call him home,” Robin Williams responded with, “Is God some man named Vinnie saying, ‘Give me my money?’”

(20) One definition of heaven is hotels run by the Germans, cooking by the French, lovers who are Italian, bankers who are Swiss and police who are English. The corresponding definition of hell is hotels run by the French, cooking by the English, lovers who are Swiss, bankers who are Italian and policemen who are German. (Tonight Show starring Johnny Carson, first week of Dec. 1990)

The main objective of the necessary semantic theory should be to establish an inventory of the types of semantic triggers which hearers (both native and non-native) use to recover information when it has been ellipsed or deleted. In further work, I will propose a system of classification, outlining the levels of and interaction between these triggers (Dissertation, in preparation). While I realize that the problem of semantic recovery is tantamount to the problem of the semantics of language itself, I will explore to what extent such a system of classification is attainable. Thus the dissertation will contribute to the study of ellipsis and information recovery as well as to the study of semantics as a whole.

Applications of Semantic Recovery Strategies

5.1 Second Language Acquisition

The issue of recoverability is a particularly important one for second language learners as they are continually engaged in these types of interpretive processes. This process of decoding and retrieval, which native speakers are incredibly adept at, is a significant acquisition in the strategic competence of the language learner.

As we shall see below, each of these processes the non-native speaker must incorporate into their interpretations, are not readily available to them, especially in the initial or even intermediate stages of language learning and acquisition. They must be taught to the non-native speaker in the language learning process.

The second language learner carries the maxims of cooperation with him or her from the native language, though, as I will argue, they are modified according to the specific L1 culture as numerous studies in contrastive rhetoric have discovered (cf. Kaplan [1966]; Eggington [1987]). In conversations with native speakers, however, the principle of cooperation contributes to the non-native speaker's sense that what is being communicated is relevant and contains enough information to be understood, but the non-native hearer still may not have enough clues to recover the ellipsed material because of their lack of linguistic skills in the L2. I shall discuss this more specifically in the dissertation.

The examples in (10-11) (reprinted)
prove particularly problematic for the non-native speaker who may wish to assume relevance, but may not understand its linguistic realization in either (10) or (11). Kato (1986) examines several types of ellipsis and discusses the difficulties non-native speakers of a language encounter in attempting to recover the ellipsed information. Some of her examples include (12)-(17):

(10) A: Where's Kevin?
 B: There's a yellow VW parked outside Sheila's house.

(11) A: Want to go to the movies this afternoon?
 B: There's a sale at Penny's.

(12) All in favor?

(13) She swallowed hard.

(14) My father was up and dressed. My mother was up, wearing a robe. My sister came running out of her room in a robe, and one by one she and I opened our packages. (William Saroyan, Papa You're Crazy)

(15) “Congressman Ryan, you are a mother—er,” Sly yelled

(16a) Grüß Dich! (I say hello to you)
 (b) Grüß Fritz! (Say hello to Fritz)
 (c) Grüß Gott! (I say hello to you [in the name of God])

(17) The woman insisted they stay for a drink. She even gestured toward Arlene, one woman to another. Arlene caught sight of the chipped polish, peachy pink. (Joyce Carol Oates, The Madwoman).

Each of the above examples has been difficult for non-native speakers to parse. Many of these have also bewildered professional translators, who were also non-native speakers, but who have a great deal more experience than the average non-native speaker attempting to understand and interpret the new language correctly. (12) is a formulaic ellipse in English which the translator recovered as:

(18) “Are you all in favor?”

rather than the more precise rendering:

(19) All [those who are] in favor [of the motion], [please signify by answering ‘aye’].

Example (13) was interpreted as “She caught her breath” while robe in (14) was translated as “a long formal dress for women (cf. robe decollete).” Kato speculates that, in the latter case, the translator/annotator could have miscued dressed as a trigger for dressed up or was unfamiliar enough with the practice of Christmas morning to use that context as the situational trigger in order to recover the correct meaning of robe. The translator of (15), taken from Newsweek magazine (4 December, 1978), interpreted the string of hyphens as he/she would in his/her own language (Japanese) to indicate the lengthening of the preceding vowel and so translated the phrase into the English equivalent of:


thus failing to recover the information intended in the original English morphemic ellipsis. Likewise, Kato argues that foreign learners of German have difficulty with the “routine” formulas represented in (16a-c).

(17) is an interesting recovery process even for native speakers, while non-native who may not have the trigger polish available to them as a clipping (ellipsis) of nail polish would not understand that Arlene could see the woman’s gesturing hands and that those hands contained a “peachy pink” nail polish which was chipping (had probably been worn awhile). This example illustrates two of the most significant problems in formulating a theory of information recovery for both native and non-native speakers: what strategies do native and non-native speakers use to recover information, and what are the boundaries for how much information to recover?

Second language learners usually resort to two interpretation strategies in order to parse texts: some type of translation method, and a word by word combinatorial aggregation to establish meaning (the latter strategy is supported by those who study reading strategies, cf. Correll & Devine, in Sometimes).

Native speakers rely highly on the sub-categorization principles within their mental lexicons (i.e., to know that the verb put takes two NP arguments, the first which takes the thematic role of patient or theme and the second which takes the role of location as in “put the book on the table”). The problem with such a strategy is obvious in that it assumes a high degree of nativization in the language in order to be able to activate information which native speakers do almost instantly. Other native strategies (as noted above), such as cohesion and coherence, entailment, anaphora, cataphora, presuppositions, local reduction, local synonymy (local in these two instances refers to the ungeneralizable reduction or establishment of a synonym for the immediate text only), paraphrase relationships, transformations, theme/theme identification, ability to recognize and manipulate phrasal constraints (ala Chomsky, 1957, Syntactic Structures), locus of Moses errors (Reder & Kusbit 1991), semantic distance, relevance judgments, allusions, mode of discourse, speech genre, governance, the ECP (empty category principle), context, ambiguity resolution, inferences, connotations, scripts (cf. Raskin 1985), metaphors and symbolism interpretation, semantic recursion, semantic redundancy, intersential context, and even extralinguistic information and world
knowledge, are all cited in the literature as ways in which native speakers recover ellipsis. Each of these, however, relies on linguistic strategies which are only acquired by the native speaker after a period of exposure ranging from several years to a lifetime. The semantic and pragmatic triggers depend on these linguistic strategies (which I will argue in my dissertation are all acting very much the same in the recovery of ellipsed information).

Some texts have tried to incorporate various strategies and evaluations into their ESL content. The book Challenges: A Process Approach to Academic English (Brown et al. 1991) contains exercises for becoming a more effective reader, including exercises in: determining vocabulary in context, learning how to scan, searching for details, and how to make inferences from the material read. The latter, a highly sophisticated linguistic ability, only achieved late in the acquisition stage, allows students to search for meaning in ways other than from the literal words of the author. From a short eight paragraph essay on a solar eclipse, students are asked to infer: the Patuk villagers’ feelings about the eclipse, what type of society exists in the village, and the intended meaning of several of the author’s phrases (p. 54). Such strategies might be commonplace in the students’ native language, but they may fail in that they do not teach the student the linguistic skills necessary to activate these semantic or even syntactic triggers.

L2 learners tend to focus on grammatical correctness; unlike native speakers, L2 language-learning students tend to derive meaning from surface structures (Brown 1987: 56; see also Silva forthcoming) while L1 learners tend to focus on deep structures. This is a significant difference in the formulation of a recovery process. With their high focus on grammar, second language students may be able to recover strict structural ellipses fairly early in the learning process, but a knowledge of the deep structure and the context of the language is necessary to recover the vast majority of the information which is ellipsed.

I believe that this work has several important implications for both language teachers and second language acquisition theory as well. I am convinced that there needs to be a greater emphasis on the teaching of these specific strategies to empower the students to both increase their proficiency in the production and understanding of texts in discourse. Investigation of this issue and related questions such as:

—What other processes or linguistic strategies will speakers use to make decisions about information recovery and retrieval?
—How do speakers decide what information to delete from their utterances to the hearers?
—What role does context play in the speakers’ deciding what to ellipses from communication?
—How does this perception change when dealing with a non-native speaker?

must be investigated in order to develop a greater understanding of this issue and its effects on language learners.

Other Applications
I will also be illustrating the benefits of these strategies in Natural Language Processing (a field which continues to demonstrate an incredible ignorance of the work in linguistics and rhetoric and composition), Humor (based on a study of 2000 jokes (Attardo et al., forthcoming), prescriptiveness and composition theory and teaching. There is a great value to making such applications in terms of the usefulness and practicality of the theory outlined in the dissertation as well as for the benefits gained through the intersection of linguistics, rhetoric and composition and cognitive theory (using reader-response theory as a departure point).

References


XIIIth International Congress of Linguists. Tokyo.


