C E N T E R F O R R E S E A R C H I N L A N G U A G E

July 1987 **Vol. 1, No. 6**

The monthly newsletter of the Center for Research in Language, University of California, San Diego, La Jolla CA 92093. *(619) 534-2536*; electronic mail: *crl@amos.ling.ucsd.edu*

CONTENTS

EDITOR'S NOTE

This is the sixth newsletter of the **CENTER FOR RESEARCH IN LANGUAGE**. **CRL** is a research center at the University of California, San Diego which unites the efforts of researchers in various disciplines, including Linguistics, Psychology, Computer Science, Artificial Intelligence, Communication, Sociology, and Philosophy, all of whom share an interest in language. We accept papers relevant to language and cognition (1 - 10 pages, preferably sent via email) and welcome response from friends and colleagues at UCSD as well as from other institutions.

Teenie Matlock, Editor, CRL Newsletter
Center for Research in Language, C-008
University of California, San Diego 92093
Telephone: *(619) 534-2536*
Email: *crl@amos.ling.ucsd.edu*

SUBSCRIPTION INFORMATION

If you are currently receiving a hardcopy of the newsletter only and have an email address to which the newsletter can be sent, please forward that information to CRL.

Can we send you the email version only to save printing and mailing costs? If you require a hardcopy in addition, please request it and we will be happy to send you one.

If you require the unformated nroff version of the newsletter, please request it from CRL after you have received the current regular formatted version.

If you know of others who would be interested in receiving the newsletter, please forward the email or postal mailing address to CRL. Thank you.

BACK ISSUES

Back issues of this newsletter are available from **CRL** in hard copy as well as soft copy form. Papers featured in previous issues include the following:

The Cognitive Perspective
Ronald W. Langacker
Department of Linguistics, UCSD
no. 3, vol. 1, March 1987

Toward Connectionist Semantics
Garrison W. Cottrell
Institute for Cognitive Science, UCSD
no. 4, vol. 1, May 1987

Where is Chomsky's Bottleneck?
S. Y. Kuroda
Department of Linguistics, UCSD
no. 5, vol. 1, June 1987

JOB ANNOUNCEMENT

The Department of Linguistics at the University of California, San Diego seeks to fill a tenure-track Assistant Professor position, beginning September 1988. The candidate should be someone who approaches the study of meaning and grammar from a cognitive and/or functional perspective. Extensive experience with one or more non-Indo-European languages is desirable. Annual salary is $29,800-$37,200. The Ph.D. in linguistics is required. Send letter of application, curriculum vitae, names of 3 referees, and 1 representative publication, to:

Cognitive Search Committee
Department of Linguistics, C-008-C
University of California, San Diego
La Jolla, CA 92093

Application materials must be received no later than December 1, 1987. The University of California is an equal opportunity, affirmative action employer.

The Department of Linguistics at the University of California, San Diego seeks to fill a tenure-track Assistant Professor position in the area of syntactic theory, beginning September 1988. Annual salary is $29,800-$37,200. The Ph.D. in linguistics is required. Send letter of application, curriculum vitae, names of 3 referees, and 1 representative publication, to:

Syntax Search Committee
Department of Linguistics, C-008-C
University of California, San Diego
La Jolla, CA 92093

Application materials must be received no later than December 1, 1987. The University of California is an equal opportunity, affirmative action employer.

Dimensions of Ambiguity

Peter Norvig
Computer Science Dept.
University of California, Berkeley

Introduction

Traditionally, there have been four main dimensions along which ambiguous sentences, phrases, and words have been classified:

- Linguistic/Referential Ambiguity
- Ambiguity/Vagueness
- Lexical/Structural Ambiguity
- Polysemy/Homonomy

The linguistic/referential dimension accounts for deictic references. The sentence *''I'll do it tomorrow''* is referentially vague, but linguistically unambiguous. A particular usage of this sentence will often make it clear who *''I''* is, when *''tomorrow''* is, and what is required to *''do it.''*

Language necessarily describes the world incompletely, and thus must leave some things vague. *''I'll do it tomorrow''* is vague as to whether the act is to be done tomorrow morning or tomorrow afternoon, but this, as [Zwicky and Sadock, 1975] point out, shouldn't count as true ambiguity. In contrast, *''I'll do it next Friday,''* when spoken on Sunday, is ambiguous between 5 and 12 days hence (at least in some dialects).

Structural ambiguity occurs when a phrase has several distinct syntactic parses. For example, *''spaghetti with meatballs and wine''* would be interpreted differently than *''spaghetti with bread and butter''* under the standard gastronomic assumptions, but both have two possible parses. This is opposed to lexical ambiguity, where the constituent analysis is constant across interpretations, but the choice of lexical units changes. For example, *''I went to the bank''* is lexically ambiguous because *bank* has at least two readings as a noun.

There is a further distinction to be made between lexical items that are *polysemes* and those that are *homonyms.* The money-bank and river-bank senses of bank seem unrelated, and are called homonyms, while the senses of, say, *''cake''* - chocolate cake, fish cake, cake of soap - are closely related and are called polysemes.

The traditional dimensions are useful in the context of a generative theory which sees a language as a set of grammatical strings, each with a set of possible interpretations. Under this view, the rules and lexicon of a generative grammar determine a mapping between strings of words and interpretations, or meanings. A string with exactly one interpretation is unambiguous, one with no interpretation is anomalous, and one with multiple interpretations is ambiguous. To enumerate the possible meanings is the proper job of a linguist; to then choose from the possibilities the one ''correct'' meaning is a problem in pragmatics, or in Artificial Intelligence.

Unfortunately, this characterization of language, and of meaning, fails to account for certain psychological facts of human language usage. It is a psychological fact that we often listen to ambiguous sentences and choose a single interpretation, without being consciously aware of considering the complete set of possibilities. In other cases, we may notice ambiguity, but have a clear preference for one interpretation over another, and in still other cases we may be genuinely confused. To compound the problem, we can change our mind about the meaning of a sentence - ''at first I thought it meant this, but actually it means that.'' Finally, our affectual reaction to ambiguity is variable. Ambiguity can be humorous, confusing, or perfectly harmonious. To a psycholinguist or AI researcher, these facts are crying out to be explained. I will attempt an explanation by investigating a variety of ambiguous sentences, and citing the problems they pose. This will lead to some conclusions about a theory of language use, understanding, and meaning.

Garden Path Sentences

A garden path sentence invites the listener to consider one possible parse, and then at the end forces him to abandon this parse in favor of another. Listeners are conscious of this switch, and often have difficulty with it. A well-known example is 1.1, where *raced* is initially treated as a past tense verb. This analysis fails when the verb *fell* is encountered; after some difficulty *raced* can be re-analyzed as a past participle. For most informants, there is a distinct feeling of having to re-parse the sentence; it does not feel like both parses were being built up simultaneously, and the second one was tested after the first was ruled out.

Most informants find 1.2 to be much less of a garden path than 1.1. An explanation for this is that while the listener is parsing *''fell down and broke its leg''* as a verb phrase, he or she is also trying to re-analyze *''the horse raced past the barn''* as a noun phrase. There is sufficient time to do this re-analysis before the end of the sentence in 1.2, but not in 1.1. Thus, 1.1 is very confusing, while 1.2 is not. We can try to quantify this analysis by asking how long does it take to re-analyze a sentence. If you agree that even 1.3 is much better than 1.1, then the answer may be as little as one word.

1.1 The horse raced past the barn fell.
1.2 The horse raced past the barn fell down and broke its leg.
1.3 The horse raced past the barn fell down.
1.4 The horse raced at the Belmont died.

Now consider 1.4, which is easier to understand than 1.1, despite having identical syntactic structure. One explanation for this is that *''raced past the barn''* is just not as good a descriptive attribute as *''raced at the Belmont.''* The first modifier could be true of any healthy horse, while the second describes only a top race horse. Thus, while both make good past tense verb phrases, only 1.4 makes a good past participle modifier.

Another example of this distinction is illustrated in 2.1 and 2.2. These examples are complex because ''got'' is highly polysemous. It can mean 'received' (as in *got a raise*), 'became' (*got old*), 'undergo' (*got arrested*) or 'cause/achieve' (*got them arrested*). In 2 the initial interpretation is 'the boy became obese', with *''fat''* interpreted as an adjective. When the final word is processed, this initial interpretation has to be abandoned.

2.1 The boy got fat spattered.
2.2 The boy got fat spattered on his arm.

An informal experiment in [Schubert, 1984] shows that 2.1 is a quite difficult garden path sentence, while 2.2 is not a garden path. One explanation for this is that having fat spattered on one's arm is the kind of experience one might be described as having undergone, while 'fat spattered' is not something one undergoes, nor is it the kind of thing one normally strives to achieve. Also, as pointed out above, 2.1 is difficult because it asks the listener to abandon the 'became obese' interpretation and come up with a new one in the course of one word.

Selection Restrictions

It seems that many ambiguous interpretations are not consciously considered because they violate selection restrictions. For example, in *''I drink port,''* the noun *''port''* is unambiguously interpreted as a kind of fortified wine, even though it also has senses meaning a harbor, and the left side of a ship. Thus, we might be tempted to formulate a principle stating that senses violating selection restrictions are not considered when there is another sense that satisfies the restrictions. Unfortunately, the sentences in 3 are evidence against this principle as it stands:

3.1 The astronomer married a star.
3.2 The plumber lit his pipe.
3.3 The rabbi was hit on the temple.
3.4 The hay farmer drank through a straw.

In each of these, the final noun has one meaning that satisfies the selection restrictions. However, there is another meaning that is suggested first, and which stubbornly refuses to go away.

Another problem is that selection restrictions are not really restrictions at all, but are more like preferences. Consider 4.1 below, which is ambiguous between the chicken being the agent and the object of eating. 4.2 prefers the agent interpretation, because dogs in our culture eat but aren't eaten. 4.3 prefers the object interpretation, because clams are eaten, and while they may eat, we are reluctant to attribute to them sufficient faculties to be in a state of mental readiness. Still, with the proper context, the preferred meaning of any of these sentences can be reversed.

4.1 The chicken is ready to eat.
4.2 The dog is ready to eat.
4.3 The clams are ready to eat.

Syntactic Preferences

[Schubert, 1984] provides a summary of previous research on syntactic reasons for preferring one interpretation over another. The principle of Right Association says that PP's tend to attach to the most recent VP or NP they could possibly modify. Thus, in 5.1, the preferred reading is that ''*for Mary*'' modifies ''*selected,*'' not ''*book*'' or ''*bought.*'' In 5.2, however, the preferred reading has the PP modifying ''*carried,*'' not ''*groceries.*'' This is explained by the principle of Minimal Attachment, which prefers parses that use the longest rewrite rules, and thus result in a parse with fewer nodes. If we assume a grammar which includes the rules VP -> V NP PP and NP -> NP PP, then attaching the PP to the V rather than the NP minimizes the number of nodes. This analysis presupposes that Minimal Attachment takes precedence over Right Association, and that ''*carried*'' subcategorizes for the VP -> V NP PP rule, while ''*bought*'' does not.

5.1 John bought the book which I had selected for Mary.
5.2 John carried the groceries for Mary.
5.3 John met the girl that he married at a dance.
5.4 John saw the bird with the powerful beak.
5.5 John met the girl that he saw at a dance.
5.6 John saw the bird with the powerful binoculars.

If we compare 5.1 with 5.3 and 5.2 with 5.4, we see that the preferences can be reversed with the proper semantic context (and can be reversed back again in 5.5 and 5.6). At best then, these syntactic preferences are only one factor that must be considered in arriving at the best interpretation.

Mutually Compatible Interpretations

Consider the following quote from Richard Parsons, of the American Fur Industry Inc., on their new advertising slogan ''Fur is for Life'': ''it has a good sound, a good connotation. Yes, they last a long time. Yes, they're a good product. Yes, furs support wildlife conservation.'' Parsons (although not a professional linguist) is making a claim about language use: that the proper or intended meaning of a phrase can be a combination of a number of interpretations, rather than a selection of one unique interpretation. In all the ambiguous phrases we have seen so far, interpretations seem to compete with one another. We can switch back and forth between two interpretations, but cannot accept both at once. This is similar to the Necker cube effect in visual perception. But Parsons is saying that the phrase ''fur is for life'' is different. Five interpretations of the slogan are listed below. Presumably, Parsons would like the public to accept 6.1-4 as mutually compatible, and rule out 6.5 as incompatible, or better yet, to never consciously consider 6.5 at all.

6.1 Fur is durable.
6.2 The fur industry is pro-conservation.
6.3 Fur wearers are lively.
6.4 The recipient of a fur may become indebted to the giver for life.
6.5 Fur, while on an animal, protects its life.

While Parson's claim is a radical departure from the 'one string/one interpretation' theory of meaning discussed in the introduction, it is in fact the norm in rhetoric, in poetry, and, it seems, in advertising (see

[Burli-Storz, 1980]). To support this claim, I opened a poetry anthology at random to Dylan Thomas, finding the opening line of his poem In the Beginning: *''In the beginning was the three-pointed star.''* As the rest of the poem makes clear, the *three-pointed star* should be taken as referring simultaneously to a stellar body in primordial space, to the light in God's performative speech act *''Let there be light''*, to the star of Bethlehem, and to the Holy Trinity. Many more examples of this type are provided in [Lakoff and Turner, 1987].

To take another example, a pop song by the Talking Heads proclaims ''We are creatures of love.'' This can be taken as having the three interpretations listed in 7. Thus, not only does the genitive have three mutually compatible interpretations, but the word *''love''* has two.

7.1 We are born as a consequence of sexual love.
7.2 We have souls that contain or are composed largely of love.
7.3 We are possessed by the force of love.

There are also cases of multiple interpretations which don't involve poetic license. Consider the use of ''book'' in ''This book, although beautifully bound, contains only one new idea in 500,000 words.'' The use of ''beautifully bound'' refers to a physical object, ''one new idea'' refers to the abstract content, and ''500,000 words'' refers to a particular instantiation of the content. (Presumably if the book were translated into another language, it would have a different number of words, but still only one new idea.) All three polysemous interpretations of ''book'' are used simultaneously. It is *not* the case that ''book'' has one meaning that encompasses these three components. This can be seen by looking at other uses where the various senses are separated out. In describing a prolific romance novelist, we can say ''Barbara has written over 50 books,'' or we can say ''Barbara has really only written one book, over and over.'' However, we cannot say ''Barbara is the author of over 7 million books,'' even if that many books with her name on the cover have indeed been printed.

Meaning vs. Connotation

One could object to some of the claims in the previous section by asserting that, say, ''three-pointed star'' has only one literal meaning, but that it has several connotations in the context of the poem it appears in. One problem with this approach is that it just postpones the inevitable – we can perhaps arrive at a complete theory of meaning, but the difficulties remain when we turn to the theory of connotation.

[Wilensky, 1987] points out several points of confusion in the literal meaning/sentence meaning dichotomy, and proposes a new dimension: the primal content/actual content distinction. Primal content is derived from lexical and grammatical knowledge, but it may not even be a ''meaning'' *per se*. On the other hand, actual content refers to the final interpretation arrived at by the listener. Wilensky's distinction seems to be compatible with the approach outlined in this paper.

Jokes and Puns

A wide variety of jokes – and all puns – rely on ambiguity. The problem is to explain why they are funny, while other ambiguities are not. Why is it that, to my ears at least, ''the rabbi was hit on the temple'' is funny, while ''the plumber lit his pipe'' is merely confusing? [Freud, 1916] claims that the laughter response is illicited by the release of supressed violent or sexual thoughts. Freud also presents the standard definition of joking as the ability to find hidden similarities between dissimilar things. This is amended to allow for the discovery of differences, or just ''to bind into a unity, with surprising rapidity, several ideas which are in fact alien to one another.'' He cites as an example the joke ''I met Baron Rothschild, and he treated me quite as his equal – quite famillionairely.'' This is funny because of the unexpected ease of combining 'familiarly' with 'millionaire' to create a new word meaning 'as familiarly as is possible for a millionaire.' Such 'innocent' word play and nonsense jokes were not explained by Freud's theory.

[Minsky, 1980] recasts Freud's notions into the terminology of mental agents acting as censors to violent or sexual thoughts. Similar censors work to detect irrational thought, and the laughter response serves to 'shake up' the mind, get it back on track, and post a warning to take such thoughts seriously. Thus, Minsky covers both tendentious jokes and innocent nonsense in the same theory.

Pragmatic Ambiguity

Consider 8 below:

8.1 Pregnant women who smoke risk premature birth.
8.2 Fetuses with mothers who smoke risk premature birth.

In both sentences, *''premature birth''* is a syntactically unambiguous noun phrase which is ambiguous on pragmatic grounds: it may be the birth *by* someone or the birth *of* someone. The former interpretation is preferred in 8.1 and the later in 8.2 because the unborn are normally not pregnant, women, or smokers.

Anomalous Strings

We have seen that it is an error to assume that strings with multiple interpretations must pick exactly one as their meaning. But what of strings with no valid parses? Chomsky has argued that strings like *''colorless green ideas''* are grammatical yet semantically anomalous. It seems there are also ungrammatical strings which can be assigned semantic interpretations. As it stands, 9.1 below has no parses, but we nevertheless understand it as a corruption of 9.2, and not of 9.3, even though either of these could be derived by changing two words in 9.1. We know from the ?/* notation that there are degrees of grammaticality, but it seems that even clearly ungrammatical strings can nevertheless have intended meanings.

9.1 John and me is running a race.
9.2 John and I are running a race.
9.3 John beat me in running a race.

Conclusions and Future Research

This paper raises many questions, but fails to answer them all. We are left with a theory of meaning where there is an important distinction between consciously considering an interpretation and eventually accepting, rejecting, or revising (correcting errors in) the interpretation. This presupposes some way of ordering interpretations as to their promise; some valid interpretations will never even be considered, because others will be looked at first. The process of arriving at a final interpretation can thus be seen as a best-first search. One complication with this view is that the final interpretation can include multiple components, not just the first one arrived at.

We are working on a text-understanding program that will account for the many dimensions of ambiguity mentioned in this paper, and will incorporate the consider/accept dichotomy for interpretation rules, as well as the primal/actual content distinction.

A preliminary version of the program has been implemented. It displays graphical representations of both the syntactic parse and the semantic feature structure of an input sentence, and allows the user tomanipulate these graphs and change the grammar and knowledge base interactively. Unfortunately, the program currently recognizes all possible interpretations, without choosing among them. Progress on this project will be reported in another paper.

Acknowledgement

I would like to acknowledge Yigal Arens for helpful comments on an earlier draft of this paper.

References

[Burli-Storz, 1980] Burli-Storz, Claudia, Deliberate Ambiguity in Advertising, 1980.

[Freud, 1916] Freud, Sigmund, Jokes and Their Relation to the Unconscious, Translated by James

Strachey, 1960, 1916.

[Lakoff and Turner, 1987] Lakoff, George and Mark Turner, More Than Cool Reason, to appear, 1987.

[Minsky, 1980] Minsky, Marvin, ‘‘Jokes and the Logic of the Cognitive Unconscious’’, AI memo No. 603, MIT AI Lab, 1980.

[Schubert, 1984] Schubert, Lenhart K., ‘‘On parsing preferences’’, *Proceedings of Coling84*, Stanford, CA, July 1984.

[Wilensky, 1987] Wilensky, Robert, ‘‘Primal Content and Actual Content: An Antidote to Literal Meaning’’, Technical Report, to appear, Computer Science Division, UC Berkeley, 1987.

[Zwicky and Sadock, 1975] Zwicky, Arnold M. and Jerrold M. Sadock, ‘‘Ambiguity Tests and How to Fail Them’’, in *Syntax and Semantics*, vol. 4 , Academic Press, NY, NY, 1975.