Inferential Aspects of the Cloze Task

Michael C. McKenna
Witchita State University

WESTERN MICHIGAN UNIVERSITY
Dorothy J. McGinnis
Reading Center and Clinic

INFERENTIAL ASPECTS OF THE CLOZE TASK

Michael C. McKenna

WITCHITA STATE UNIVERSITY

The central assumption underlying the cloze procedure is that context can be used inferentially to predict deleted words (or other graphic units). The importance of this assumption suggests that an adequate understanding of the cloze task is impossible without an analytical knowledge of the role inference must play. The present discussion embodies a stepwise analysis of this role, from the point at which the context is read to the point at which the reader is able to distinguish acceptable responses.

1. Inferring Constraints

An intriguing question common to logic and to reading is the classification and explication of inferential thought. Within the field of reading, efforts have been made to identify the types of information which make inferences possible about the meaning of an unfamiliar word (Artley, 1943; McCullough, 1958; Ames, 1965). The case of cloze deletions is analogous in that subjects must draw upon the same sorts of clues in an attempt to constrain the meaning of the missing word, and thus the number of alternatives, as much as possible.

For present purposes, it is unnecessary to enumerate these types, and, to repeat, such an effort at this writing would have to be an incomplete one.[1] It is enough here to make three generalizations concerning the process of inferring constraints.

First, it shall be assumed that constraints can be inferred from context in such a way that they are countable. That is to say, from the words which surround a deletion, it is possible to enumerate distinct conclusions about the meaning of the deleted word. That such constraints are interrelated will be seen presently.

Second, the inferred constraints are of two varieties: *syntactic* (limitations upon the grammatical class, or "part of speech," of the missing word) and *semantic* (limitations upon the meaning of the missing word).

Third, semantic constraints can, and for present purposes must, be further delineated into *coordinate* and *subordinate* constraints. Subordinate constraints are those which, while identifiable as distinct "conclusions," are automatically combined by the subject because the com-

---

[1] It should be noted, however, that one entire class of clues is denied the cloze subject those found within the word itself, such as its etymology, inflection, and configuration. Thus, his circumstance is only partially analogous to that of the reader encountering unfamiliar words. For a discussion of differences between recognition and recall, see Glanzer and Bowles (1976).

bination is thought of in a singular sense (stored in association). For example, in the sentence "The __ injured one of its four legs," it is possible to infer (1) that the deleted word names an animal and (2) that the animal has four legs. But are there two constraints here or one? Since the two characteristics above are typically thought of together as determining the set of four-legged animals, they will be considered here to be subordinate constraints — combined to form a single coordinate constraint. Coordinate constraints, on the other hand, can be combined logically but are not customarily linked. For example, in "His mother named him __ because she liked monosyllables," two semantic constraints are present: (1) the word must be a boy's name and (2) it must consist of one syllable. Since words with both of these characteristics are not ordinarily thought of together, these are coordinate rather than subordinate constraints.

2. Accepting Inferred Constraints

The ability to infer constraints is not the sole component of the inferential stage of the cloze task. Subsequently, the subject must decide whether to accept a constraint in the process of identifying a set of alternative responses. For at least three reasons, he may choose to disregard some conclusion he has reached, even though it may have a demonstrable basis in the text. First, one inference may duplicate another and thus contribute nothing to the accumulation of constraint. This state of events is not limited to the pronounced redundancy of inefficient writing but occurs frequently when context is viewed at full scale. Ramanauskas (1972a, 1972b) has demonstrated that even for poor readers constraints operate between, as well as within, sentences, and it is natural to assume that many of these are duplicated as short range. Second, and in the same vein, one constraint may *imply* another without precisely duplicating it. In the sentence, "He lives in Europe, in the French city of __," the phrase "in Europe" is rendered useless by the narrower constraint inferable from "the French city of." Third, an inference may be viewed by the reader as nonessential but as possible nevertheless. He must decide whether to honor it and thus whether to place an added restriction on his target set. In so doing, he may attempt to estimate the probability that the author's meaning was accordingly constrained. But only a knowledge of the exact word can settle the issue. In the sentence, "The woman carried her __ outdoors," it is arguable that the missing word names an object customarily found out-of-doors, but the argument can never be made conclusive. In fact, the reverse can be contended with equal vigor. The problem is that the cloze deletion has created an artificial ambiguity which causes the subject to demand too much of the remaining text. He will often be aided by the broader context, but not always.

In sum, it is not enough to infer constraints. The cloze subject must then judge the strength and distinctness of his inferences and to decide accordingly whether to lend them credence. And yet even at this point the inferential process is not complete.

3. *Combining Accepted Constraints*

Once the subject has selected the constraints which bear upon the meaning of the deleted word, he must combine them. In so doing, he establishes a single category of words having several common attributes from several categories each defined by a single attribute. The process is one of concept building. All concepts are characterized by attributes which serve to define examples of the concept, and as the list of required attributes (constraints) is increased, the number of examples which possess all of them diminishes. Thus, the value of individual constraints in the cloze task becomes clear: the application of each new constraint operates to reduce successively the number of acceptable word replacements (i.e., the size of the target set). The subject may first infer that the missing word is a noun. This is a major logical step in that the number of logical choices is reduced by over 50 percent (French, Carter, & Koenig, 1930; Fries, 1952). He may further infer that the word names an inanimate object and thus again reduce the target set. The more constraints he can combine in this fashion, the better his chances of success. At this point, the importance of the coordinate-subordinate distinction becomes clear in that subordinate constraints have been combined *a priori*.
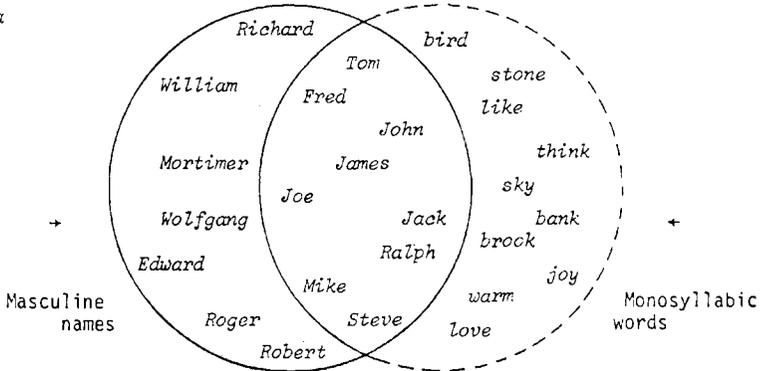
The important realization is that conclusions inferred regarding the identity of the deleted word *can* be combined. The truth of this assertion follows logically from category theory and mathematically from axiomatic set theory. There are two perspectives possible on the manner of combining constraints, and they differ in terms of the degree to which one considers the constraints independent of one another. It will be seen that either perspective leads to the identification of the same target set.

First, two constraints may be looked on as defining two sets of words which "overlap," such that there is an intersection of words common to both sets. The words found in the intersection share two attributes by definition — one required for membership in one of the sets and one for membership in the other. On the other hand, the words found in the set union, but not in the intersection, have only one or the other of these attributes. In the earlier sentence, "His mother named him __ because she liked monosyllables," two constraints are inferable, and, accordingly, the set of alternatives which meet both of them can be represented as the intersection of the sets of words which meet at least one. The intersection is, then, the set of all monosyllabic masculine names, and it is the target set for the deletion. [2]

Second, one constraint may be looked on as defining a *proper* subset of the set determined by another constraint. That is, the set of words delimited by one constraint may be seen to contain *only* words previously identified by another constraint. One begins with a single inference and its word-

---

[2] Owing to individual differences in memory content and passage interpretation, the target set will actually vary with the subject. While logic may occasionally make possible certain "absolute" pronouncements like the one above, the matter is in reality much more complex.

set – say, the set of all masculine names – and uses the second constraint to locate a subset of the first viz., those masculine names which happen to consist of one syllable. No consideration is given to monosyllabic words which are not masculine names, as it was in the previous case. The contrast between these viewpoints is portrayed in Figure a, where elements bounded by the broken line are considered only from the first perspective, in which the intersection of two sets is seen rather than a subset of one.

a

Richard Tom bird
William Fred stone
like
John
Mortimer James think
Joe sky
Wolfgang Jack bank
Ralph brock
Edward
Mike joy
Masculine warm
names Roger Steve love
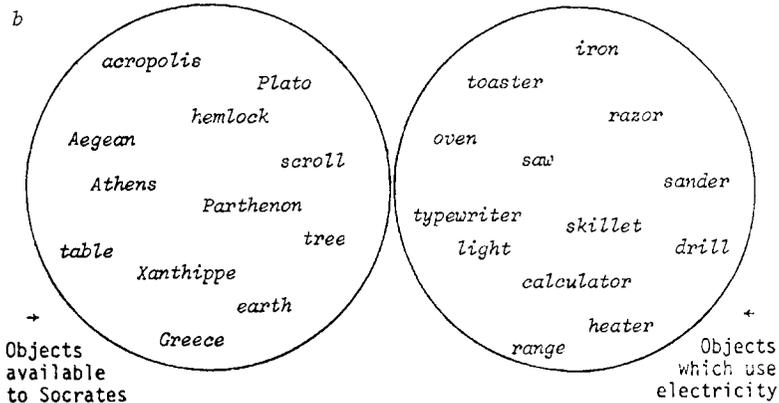Robert

→ Masculine names ← Monosyllabic words

His mother named him _____ because she liked monosyllables.

It is difficult to imagine a set of *functional* constraints which can be viewed in only one of these ways. When one constraint is implied by another, as in the example of the French city, it is, of course, only possible to view their combination from the second standpoint above (since there are no French cities which are not also in Europe). But since only the narrow constraint is functional, it is assumed that the reader will discard the first rather than attempt a formal (and pointless) combination. The important principle is that the target set which is ultimately defined will be the same regardless of how the combinative process is viewed.

Insofar as the *nature* of the target set is concerned, an important variant remains to be considered. Actually, it rarely, if ever, occurs in rational prose, but its implications are far-reaching. This is the case in which two or more constraints define sets of words without a mutual intersection. That is to say, no word exists which possesses the attribute required by each individual constraint. From the sentence, "Socrates plugged in his electric ___," two inferences follow which are seen to be contradictory. The deleted word names an object available to Socrates, and it also names an object which makes use of alternating current. Since nothing possesses both of these attributes, the sets of words corresponding to them are disjoint (Figure b). While this circumstance is not historically acceptable in normal

discourse, it serves to describe what at times is an analogous situation in the mind of the cloze subject. If he is aware of the constraints and simply cannot generate a word from what he well knows to be the target set, his position is no different from the one above. The foreigner who reads "He lives in __ Diego" may suspect that the missing word is part of a place name ending with *Diego*, but his lack of familiarity with American geography may prevent his being able to generate the one-element target set, (San). [3] J. Anderson (1972) found that the cloze scores of subjects in the process of learning English as a second language are not significantly improved when synonyms are counted correct. This result suggests the economy with which new speakers acquire words— they cannot readily afford the luxury of many alternate terms. When a word must possess a number of attributes, there may simply be none available to the foreign subject whose vocabulary is impoverished to begin with. It is reasonable to suppose that vocabulary deficiencies in native speakers can likewise lead to the production of an empty target set.

b

acropolis, Plato, hemlock, Aegean, scroll, Athens, Parthenon, table, tree, Xanthippe, earth, Greece

iron, toaster, razor, oven, saw, sander, typewriter, skillet, light, drill, calculator, heater, range

→ Objects available to Socrates

← Objects which use electricity

Socrates plugged in his electric _____.

But without regard to such limitations in the reader, it may be seen that in normal writing constraints tend to be complementary rather than contradictory and are combined to define nonempty target sets. The last example serves to illustrate the possibility that such a set may contain only one element. Without involving a set-theoretic analysis, Taylor and Waldman (1970) designed such items as "unique." It is not true, however,

---

[3] It is interesting to note that there *are* two constraints operating in this sentence. The word *Diego* does not of necessity imply the place name; consider "Zorro's real name was __ Diego," the deleted word being *Don*.

that any logical difference exists in how the target set is defined. The fact that it contains only one word is a quantitative distinction, not a qualitative one.

4. *Determining the Target Set*

To identify and combine the characteristics necessary for set membership are sufficient to define a set. But to determine a target set from linguistically inferred constraints does not follow quite so directly. The problem is one of imprecision. The inferred attributes required of all words in the target set are apt to be vague or relative. In some cases, these attributes are sufficiently precise to test any candidate word for set membership, but in many others they are not. In the earlier example, "The woman carried her __ outdoors," it is reasonable to infer that the target set must include only words which name objects conveyable by a woman. But although a valid inference, the resulting constraint is imprecise as a defining characteristic. For example, would *desk* be excluded from the target set? Clearly, there are certain women capable of carrying certain desks. The issue is relative to the weight of the desk and the strength of the woman. Relative characteristics were a particular problem for Aristotle in his own attempt to explain categorization (see *Categoriae*), and they appear to be a difficulty inherent in language. Lakoff (1972) has re-emphasized the matter of what he terms "fuzzy" categories, and he has especially noted the tendency of relative adjectives to result in fuzziness — e.g., the set of all large men.

For the cloze subject, this phenomenon can be turned to an advantage. By broadly interpreting "fuzzy" constraints, he is able to increase the size of the target set at such times as he may have difficulty generating a word meeting all of the constraints (i.e., as he initially conceived them). Conversely, by narrowly defining "fuzzy" constraints, he is able to restrict the size of a target set which offers many alternatives and at the same time guarantee a "safe" choice.

Of course, this sort of adjustment is inconceivable with the precise defining characteristics required in mathematics or in logic. But the cloze subject is not often confined by such precision. Byerly (1973) has underscored the advantage one has in the ability to interpret flexibly constraints which are linguistically vague:

> A set is well defined only when we can determine whether any definite entity is or is not a member of the set. In natural language vague indications may serve a purpose. It is not always possible to draw distinctions precisely. Nor is it always convenient, as when we tell someone to "put the package down somewhere over there." (p. 329)

On the other hand, it is not difficult to see how vague constraints can pose problems, especially time problems, for the cloze subject. It is enough at this point to observe the possibility of ill-defined target sets and the fact

that cloze subjects must somehow deal with them. At stake is the crucial distinction between target and nontarget words, and it may be that a weakness in boundary contributes to the strength of the idea that subjects search within memory a "target" set which includes words subsequently deemed unsuitable (see Kaplan, Carvellas, & Metlay, 1971; Tuinman, 1972; McKenna, 1977a).

The cloze subject's ultimate goal of generating the one word actually deleted requires that he first establish target set boundaries which correspond to those conceptualized by the author when he selected the word. The relationship can therefore be studied from a set-theoretic perspective, from which it can be generalized that the overlap of the two target sets (i.e., writer's and reader's) is one determiner of the subject's success. McKenna (1977b) used a set-theoretic approach to study the relationships among intended, expressed, and reconstructed meaning in the writer-to-reader sequence. It would appear that the cloze task is a special case of those general considerations.

5. *Summary*

On the basis of context, inferences must be made, accepted, and combined concerning the meaning of the missing word. The strength of such conclusions must be judged and limits set as to the probability and suitability of individual words. All of this is not to say, however, that these steps occur in a rigid sequence or that they proceed at all times under the conscious direction of the subject. The attempt has been to describe certain events which must take place, at least in effect, if a successful response is to be generated. It is important to observe that simply because inferential thought is not *verbalized* (as it has been in these pages) does not imply its absence.

REFERENCES

Ames, W. A study of the process by which readers determine word meaning through the use of context (Doctoral dissertation, University of Missouri, 1965). *Dissertation Abstracts International*, 1966, *26*, 6539-6540. (University Microfilms No. 65-14, 463)

Anderson, J. The application of cloze procedure to English learned as a foreign language in Papua and New Guinea. *English Language Teaching Journal*, 1972, *27*, 66-72.

Aristotle. *Categoriae*. In W. D. Ross (Ed. and trans.), *The works of Aristotle translated into English under the editorship of W. D. Ross* (Vol. 1). Oxford: Clarendon Press, 1908.

Artley, A. S. Teaching word meaning through context. *Elementary English Review*, 1943, *20*, 68-74.

Byerly, H. C. *A primer of logic*. New York: Harper & Row, 1973.

French, N. R., Carter, C. W., & Koenig, W. The words and sounds of telephone conversations. *Bell System Technical Journal*, 1930, *9*, 290-324.

Fries, C. C. *The structure of English*. New York: Harcourt, Brace, & Co., 1952.

Glanzer, M., & Bowles, N. Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1976, *2*, 21-31.

Kaplan, I. T. Carvellas, T., & Metlay, W. Effect of context on verbal recall. *Journal of Verbal Learning and Verbal Behavior*, 1971, *10*, 207-212.

Lakoff, G. Hedges: A study in measuring criteria and the logic of fuzzy concepts. *Papers from the eighth regional meeting, Chicago Linguistics Society*. Chicago: University of Chicago Linguistics Department, 1972.

McCullough, C. M. Context aids in reading. *Reading Teacher*, 1958, *11*, 225-229.

McKenna, M. C. An investigation of the cloze procedure as a memory-search process (Doctoral dissertation, University of Missouri, 1977). *Dissertation Abstracts International* (in press) (a)

McKenna, M. C. Reading viewed as the result of writing. *Reading Horizons*, 1977, *17*, 264-267. (b)

Ramanauskas, S. Contextual constraints beyond a sentence on cloze responses of mentally retarded children. *American Journal of Mental Deficiency*, 1972, *77*, 338-345. (a)

Ramanauskas, S. Contextual constraints beyond a sentence on cloze responses of cloze readability measures to linguistic variables operating over segments of text longer than a sentence. *Reading Research Quarterly*, 1972, *8*, 72-91. (b)

Taylor, W. L., & Waldman, I. N. Latency and focus methods of cloze quantification. In G. B. Schick & M. M. May (Eds.), *Nineteenth yearbook of the National Reading Conference* (Vol. 2). Milwaukee: NRC, 1970.

Tuinman, J. J. Speculations on cloze as a search procedure. In F. P. Greene (Ed.), *Twenty-first yearbook of the National Reading Conference* (Vol. 1). Milwaukee, NRC, 1972.