[*Mechanical Translation*, vol.2 no.2, November 1955; pp.39-46]

# *An experimental study of ambiguity and context**

**Abraham Kaplan, Department of Philosophy, University of California, Los Angeles**

Ambiguity is the common cold of the pathology of language. The logician recognizes equivocation as a frequent source of fallacious reasoning. The student of propaganda and public opinion sees in ambiguity an enormous obstacle to successful communication. Even the sciences are not altogether free of verbalistic disputes that turn on confused multiple meanings of key terms.

Special importance attaches to ambiguity as a result of the growing interest in the possibilities of mass translation: rapid and routine translation of large bodies of material. The simplest expedient, as a first approximation, is word by word translation — a word for word substitution carried out by essentially clerical methods, very possibly by machine. But word for word substitution is hardly usable when the words of both languages are even moderately ambiguous.

It is a familiar fact that ambiguity of isolated words is reduced by the contexts of their occurrence. The total behavioral situation in which language functions is decisive in determining what will be communicated. For many problems, however, (and in particular, that of mass translation), the behavioral situation is not accessible. The 'context' (itself an ambiguous word) must here be taken to consist of the verbal setting in which the word to be interpreted occurs, i.e., the other words with which it is being used.

The problem of this study is to determine to what extent and in what ways verbal setting reduces ambiguity. Is ambiguity primarily a feature of words in isolation, or does it persist to some extent even in context? What part of the context is most effective in reducing ambiguity — for instance, how is the ambiguity of a selected word affected by the words immediately preceding and following it, as compared with the effect of the entire sentence in which it occurs? Does it matter whether the immediate context consists solely of particles ? How is the reduction in ambiguity affected by the linguistic sensitivity of the translator? By the multiplicity of senses of the isolated word? By the clarity of the word; that is, the ease with which its multiple senses are identified? These are the questions to which this study is addressed.

*Reprinted with permission of the Rand Corporation from their report P18, dated November 30, 1950, which has been out of print for several years.

Two important restrictions on this study are to be noted.

In the first place, it deals with ambiguity of single words, not homonyms (word types, not word tokens[1]): the four letters "blow" actually may constitute a single word, semantically and grammatically speaking, or may be one of several homonyms — a) to send forth a current of air, b) a wind or gale, c) a blossoming or blooming, or d) a forcible act or effort. There is no doubt that the setting usually allows us to distinguish nouns from verbs, for example, hence among homonyms which are different parts of speech. The problem here will be to distinguish the multiple senses of a single word. For instance, the verb "blow" has several senses: a) producing a noise by blowing, b) panting or puffing, c) talking loudly or boastfully, and so on. These are related senses, and as a group quite distinct from the senses of the homonym "blow" which means "to blossom." The ambiguity with which this study is concerned is thus more subtle than homonymy. Whatever analysis is to be given of the distinction between homonyms and single words, it is reasonable to suppose that the effect of context on homonym-ambiguity is more marked than that of the single-word-ambiguity here dealt with.

A second restriction on the study is this. It is not concerned with what ambiguity actually occurs in written material. The attempt is to determine the reduction of ambiguity by context, and not the actual frequencies with which ambiguities and their reductions occur. To be sure, the material selected is presumed to be sufficiently representative of actual discourse to make the results of practical relevance. But this presumption is not itself being tested here. All the cases studied are actual cases; the contexts were selected from published texts and were not constructed for the study. Nor were words selected on the basis of the kinds of contexts in which they occurred, except for certain formal requirements described below.

## Procedure

A group of "translators" was presented with a set of words, each with a number of possible meanings to be judged applicable or not. The words were first presented in isolation, then in certain standard contexts.

1 For a discussion of this distinction, and a comprehensive survey of contemporary semantics, see C. W. Morris, Signs, Language, and Behavior, 1946.

The sample was derived entirely from the literature of pure and applied mathematics. This selection was made partly because of the background of the translators used in the experiment, partly because it is commonly supposed that such material involves less ambiguity than non-scientific writing, or even that of some other scientific disciplines. The specific books used are as follows:

| | No. of Samples |
|---|---|
| Alexander, J., Colloid Chemistry. Vol. III, Chemical Catalog Co., 1931 | 15 |
| Holmboe J. et al., Dynamic Meteorology, Wiley, 1945 | 15 |
| Lefschetz S., Introduction to Topology Princeton, 1949 | 9 |
| Moulton, F. R., Introduction to Celestial Mechanics, Macmillan, 1914 | 15 |
| v. Neumann J and Morgenstern,O., Theory of Games and Economic Behavior, Princeton, 1947 | 15 |
| Richter W., Fundamentals of Industrial Electronic Circuits, McGraw Hill, 1947 | 15 |
| Stuhlman O., Introduction to Biophysics, Wiley, 1948 | 14 |
| Weyl H., Philosophy of Mathematics and Natural Science, Princeton,1949 | 12 |
| Williams C.D. and Harris E. C., Structural Design in Metals, Ronald Press, 1949 | 15 |
| Zemansky, M. W., Heat and Thermodynamics, McGraw Hill, 1943 | 15 |
| Total | 140 |

The contexts were provided by sentences selected at random from these books, not drawn, for example, solely from prosy introductory chapters. On the other hand, "symbol-heavy" sentences which would require either specialized knowledge or considerable portions of text for their interpretation were omitted. Sentences were selected to vary in length from 15 to 40 words; occasionally, dependent clauses irrelevant to the clause in which the key word occurred were omitted. The distribution of sentence lengths was:

| Number of Words | Number of Sentences |
|---|---|
| 15 - 19 | 33 |
| 20 - 24 | 56 |
| 25 - 29 | 39 |
| 30 - 34 | 8 |
| 35 - 39 | 4 |
| Total | 140 |

The key words selected were limited to nouns, verbs, and adjectives; these are the major carriers of the content of any discourse, and probably more markedly exhibit ambiguities. The position of the word in the sentence was varied at random, to avoid overemphasis on the special contexts constituted by opening and closing phrases. The first and last two words of the sentence were never selected, so that contexts could be restricted to a single sentence. No mark of punctuation was allowed to occur within two words on each side of the key word, so as to simplify the appraisal of the effect of verbal setting. Only words of sufficiently general use to be included in the Fifth Edition of Webster's Collegiate Dictionary were chosen; and it was required that the dictionary distinguish at least three senses of the word.

Although frequency of use was not a criterion of selection, it was afterwards found that all of the 140 words selected appear in The Teacher's Word-Book of 30,000 Words.[2] Seventy-four of the words are among the thousand most frequent words in the English language; of these, forty-four are among the first 500. The following is the frequency of occurrence per million words in the Thorndike-Lorge count:

| Frequency | Number of cases |
|---|---|
| Over 100 | 76 |
| 50 - 99 | 31 |
| 25 - 49 | 18 |
| 2 - 24 | 15 |
| Total | 140 |

The actual key words used in the sample are listed in Table I.

For each word, a number of possible senses was listed, obtained from the dictionary entry for that word. The fully inflected form of the word was used — e.g., the plural or past tense if this was the form of its occurrence. It was required that the senses listed be clearly distinguishable (in the judgment of the experimenter) from one another; this did not by any means coincide with the numbered senses in the dictionary entry. Obsolete, archaic, colloquial, and highly technical senses were omitted. A maximum of ten senses was selected. Wherever necessary, the total number of senses was made up to ten by adding an appropriate num-

2 By E. L. Thorndike and I. Lorge, Columbia University Press, 1944.

TABLE I

Key Words Used

| | | | |
|---|---|---|---|
| appear | direct | narrow | scale |
| approaches | dropped | nature | screen |
| assume | due | new | separated |
| attached | elements | normal | serve |
| balance | established | note | set |
| bears | eye | numbers | shank |
| broad | field | observed | shape |
| care | flow | origin | show |
| case | force | part | skin |
| cells | formal | particle | slight |
| change | found | passes | solution |
| character | free | people | spirit |
| class | function | period | spread |
| classical | general | phase | state |
| clear | generation | place | strong |
| close | given | point | study |
| come | goes | position | subject |
| compose | good | possesses | substance |
| conceived | ground | power | survey |
| conditions | heads | produce | system |
| connections | heat | product | tension |
| consideration | induced | projection | terms |
| contain | introduced | properties | tests |
| contracts | leading | protection | time |
| converted | levels | provides | tool |
| course | lies | put | transmitting |
| current | little | raised | treated |
| cycle | load | reached | tubes |
| deductions | lower | reaction | types |
| degree | maintained | reference | used |
| depending | make | relations | value |
| determined | mass | requires | view |
| developed | material | rest | words |
| device | model | rise | work |
| diaphram | motion | runs | world |

ber of "false" senses, obtained from dictionary entries for words of the same part of speech. The average number of "correct" senses of the words in the sample was 5.6, approximately the degree of ambiguity in actual discourse.[3] The distribution of words in the sample with various numbers of senses was:

| Number of Senses | Number of Words |
|---|---|
| 3 | 16 |
| 4 | 33 |
| 5 | 30 |
| 6 | 25 |
| 7 | 7 |
| 8 | 14 |
| 9 | 5 |
| 10 | 10 |
| Total | 140 |

3 See G. K. Zipf, Human Behavior and the Principle of Least Effort, Addison-Wesley Press, 1949, p. 30.

Examples of words with the senses listed (including the "false" ones) are given in Table II, below.

The study was carried out with the help of seven "translators", four of whom had considerable training in the mathematical sciences, the other three having only a high school education.

Words were first presented in isolation — the so-called null context. Each translator indicated which of the ten senses for each word appeared to him to be senses in which the word might sometimes be used. In the second phase, seven contexts were employed, derived from the sentence of the actual occurrence of the word. These contexts were:

the word preceding (P1)
the word following (Fl)
both of these (Bl)
the two words preceding (P2)
the two words following (F2)
both of these (B2)
the entire sentence (S)

TABLE II

Examples of Words and Senses

Starred senses are actual ones. (Of course, no stars were printed in the sheets from which the translators worked.)

appear

1) shine faintly
*2) be obvious or manifest
*3) come before the public
4) come or go near
5) be in great plenty
*6) attend before a tribunal
*7) seem, look
8) pass or move suddenly or quickly
*9) become visible
10) look steadfastly; meditate

approaches

*1) approximations
*2) preliminary steps
3) summaries, epitomes
4) suppressions, suspensions
5) wants, lacks
*6) ways, passages
7) posterior sections
8) dwellings, sojourns
9) skills
*10) advances

assume

1) snatch, seize
2) derived by reasoning or implication
*3) suppose
4) come into possession of
*5) undertake
*6) appropriate, usurp
*7) feign, sham
8) swallow eagerly
9) hold in possession or control
*10) receive, adopt

Words were presented to the translators in one or another of these contexts, and acceptable senses were again indicated by them. The design used had the properties that each translator was presented with all the words in some context or other; each word appeared in all the contexts; each context had all the words in it; and no person faced the same word in more than one context. Thus each subject made two interpretations of each word: once in the null context, and once in some verbal setting.

Results

The accuracy of a translator was measured by the number of his correct characterizations of a listed sense as actually belonging to the word or not: ascriptions of true senses plus denials of false senses. (This measure could be used only for the null context, where the true senses are specified by the dictionary; no such standard is available for occurrences in context.) The seven translators ranged in mean accuracy for all the words from 62% to 84%, around a mean of 75%. The four trained in mathematics averaged 80% accuracy, the other three 70%. Since the isolated words are not distinctively mathematical, the difference is presumably due to general linguistic facility.

The clarity of a word is defined as the mean accuracy attained on it by the seven translators. (Like accuracy, therefore, it applies only to the null context.) The mean clarity for all the words words was 75% (being linked to the mean accuracy). The distribution was:

| Clarity (%) | No. of cases |
|---|---|
| 40-49 | 1 |
| 50-59 | 4 |
| 60 - 69 | 29 |
| 70 - 79 | 57 |
| 80 - 89 | 41 |
| 90 - 99 | 8 |
| Total | 140 |

Unclarity was not due markedly either to a failure to recognize true senses or to a tendency to ascribe false ones. The mean number of true senses was 5.6; of assigned senses, whether true or false, 5.5. Clarity did not show any significant correlation with ambiguity: words with a large number of true senses were, on the whole, neither more nor less clear than those with a small number. Neither was clarity correlated with familiarity, as measured by frequency in the Thorndike-Lorge count. In both cases the correlation was + .1 and not significant.

By the reduction of a context will be meant the ratio of the number of senses assigned to a word occurring in that context to the number assigned to it in the null context by the same translator. The lower this ratio, the more effective is the context in reducing ambiguity. The reduction of the contexts tested was found to be:

| Context | Reduction (%) |
|---|---|
| P1 | 75 |
| F1 | 57 |
| B1 | 47 |
| P2 | 50 |
| F2 | 56 |
| B2 | 44 |
| S | 47 |

The context consisting of one preceding word appears to be least effective in reducing ambiguity, being significantly worse than one word following. One word on each side of the word to be translated is more effective than two preceding or two following. It is noteworthy that two words on each side of the key word are comparable in effect to the entire sentence. The distribution of the various degrees of reduction for each of the contexts is given in the following table.

| Reduction (%) | Percent in Context | | | | | | |
|---|---|---|---|---|---|---|---|
| | P1 | F1 | B1 | P2 | F2 | B2 | S |
| 0 - 29 | 37 | 41 | 41 | 38 | 36 | 51 | 60 |
| 30 - 59 | 19 | 25 | 28 | 28 | 27 | 27 | 24 |
| 60 - 89 | 18 | 14 | 17 | 18 | 22 | 6 | 4 |
| 99 - 100 | 11 | 9 | 9 | 10 | 4 | 6 | 4 |
| over 100 | 15 | 11 | 5 | 6 | 11 | 10 | 8 |
| Total | 100 | 100 | 100 | 100 | 100 | 100 | 100 |

What is the effect of initial ambiguity on its reduction? Do more ambiguous words profit more from context than less ambiguous ones? To answer this question, words of from three to five true senses were separated from those of six to ten: there were 79 cases in the former group, 61 in the latter. The reduction effected by each context for these two groups of words was found to be:

| Context | Reduction (%) for less ambiguous words | Reduction (%) for more ambiguous words |
|---|---|---|
| P1 | 65 | 88 |
| F1 | 62 | 51 |
| B1 | 48 | 45 |
| P2 | 56 | 43 |
| F2 | 52 | 61 |
| B2 | 44 | 44 |
| S | 47 | 47 |

As can be seen, there was no consistent direction of difference: the mean reduction was 53.4% for the less ambiguous words, 54.1% for the more ambiguous. It is to be noted that P1 again appears as the worst context; B1 as quite good, and B2 comparable in effect to that of the entire sentence.

The same procedure was used to appraise the effect of clarity on reduction of ambiguity. The sample was evenly divided into words of relatively high and low clarity, as defined above, and reduction separately computed:

| Context | Reduction (%) for clear words | Reduction (%) for unclear words |
|---|---|---|
| P1 | 88 | 62 |
| F1 | 53 | 62 |
| B1 | 47 | 47 |
| P2 | 49 | 52 |
| F2 | 5? | 59 |
| B2 | 48 | 41 |
| S | 58 | 36 |

The effect is again not a consistent one, though it suggests some slight advantage to the initially unclear words, as profiting more from context. The mean reduction was 56.6% for the clear words, and 51.3% for the unclear.

The effect of familiarity was appraised in the same way. The seventy-four words which, according to the Thorndike-Lorge count, are among the thousand most frequent in the English language were separated from the remaining sixty-six words in the sample, and reduction again separately computed:

| Context | Reduction (%) for frequent words | Reduction (%) for infrequent words |
|---|---|---|
| P1 | 89 | 59 |
| F1 | 56 | 59 |
| B1 | 49 | 44 |
| P2 | 40 | 62 |
| F2 | 59 | 52 |
| B2 | 44 | 45 |
| S | 51 | 43 |

Again there is no consistent effect, though again there is some slight advantage for the less frequently appearing words, their mean reduction being 52.0% as compared with 55.4% for the more frequent ones. It is quite in accord with expectation, of course, that the less clear, less familiar words should profit more by being put in context than those that are clear and familiar to start with. But the results can only be said to be compatible with this expectation, and scarcely to confirm it.

By contrast with these slight effects of doubtful significance are two other factors which appear to be quite important in reducing ambiguity. The first is the semantic content of the context. A context might consist entirely of articles, prepositions, conjunctions, etc., and could be expected to contribute less to a translation than one which also contained words not so poor in semantic content. We may call the first particle contexts, the second substantive contexts. A context was classified as "substantive" if at least one word in it was not a "particle" word. The full list of words in the sample regarded as "particles" (not grammatically, but from the viewpoint of semantic content) is given in Table III, below. The results were the following:

| Type of Context | Particle Contexts | | Substantive Contexts | |
|---|---|---|---|---|
| | No. Cases | Reduction (%) | No. Cases | Reduction (%) |
| P1 | 89 | 80 | 51 | 66 |
| F1 | 107 | 66 | 33 | 28 |
| B1 | 67 | 54 | 73 | 40 |
| P2 | 56 | 61 | 84 | 43 |
| F2 | 62 | 62 | 78 | 51 |
| B2 | 25 | 45 | 115 | 44 |
| S | 0 | — | 140 | 47 |

---

The effect is consistent and unmistakable. The mean reduction for the particle contexts was 61.3%, for the substantive contexts, 45.6%. How effective a context is in reducing ambiguity is a function, therefore, of whether it itself has a semantic content or is functioning primarily syntactically. It is noteworthy that for the B2 context there was no significant difference in reduction; but the small number of cases of B2 particle contexts (25) makes this result suspect.

A second markedly significant factor in reduction of ambiguity by context is the accuracy of the translators. The samples translated by the three most accurate and those by the three least accurate (for the words which they were each interpreting in the context in question) were grouped separately, there being sixty cases for each group. The results were:

| Context | Reduction (%) for inaccurate translators | Reduction (%) for accurate translators |
|---|---|---|
| P1 | 109 | 59 |
| F1 | 67 | 51 |
| B1 | 58 | 46 |
| P2 | 57 | 48 |
| F2 | 63 | 52 |
| B2 | 60 | 36 |
| S | 76 | 26 |

TABLE III

List of "Particles"

| | | | |
|---|---|---|---|
| a | from | only | they |
| above | has | or | this |
| against | if | other | through |
| all | in | our | thus |
| an | into | out | to |
| and | is | over | under |
| are | it | quite | until |
| as | its | same | us |
| at | just | several | very |
| be | let | shall | we |
| behind | many | since | when |
| between | may | so | which |
| by | must | some | whose |
| can | near | than | will |
| certain | no | that | with |
| does | not | the | within |
| done | of | their | would |
| during | on | there | |
| for | one | these | |

The effect is again unmistakable. The inaccurate translators showed a mean reduction, for the various contexts, of 70.0%, while the accurate translators attained a reduction of 45.5%. In the sentential context, the reduction of the accurate group was about three times as great as that of the inaccurate group.

In terms of these two important factors, an appraisal can be made of the optimal reduction of ambiguity by context, considering only the accurate translators, working with substantive contexts. The results are:

| Context | No. Cases | Reduction (%) |
|---|---|---|
| P1 | 24 | 40 |
| F1 | 13 | 35 |
| B1 | 35 | 33 |
| P2 | 38 | 39 |
| F2 | 29 | 42 |
| B2 | 53 | 36 |
| S | 60 | 26 |

Conclusions

1. Even for familiar words, no more than about 3/4 of the possible meanings presented are correctly translated as senses in which the words might sometimes be used.

2. The accuracy of such translation varies significantly from person to person, and shows some relation to educational level. Whether this is due to language ability, intelligence, or some other factor was not investigated.

3. There is no consistent direction of error in translation: false senses are as likely to be ascribed to words as are true senses to be unrecognized,

4. How accurately, on the whole, a word is translated bears no marked relation to the number of its actual senses nor to the frequency (within a fairly wide range) of its occurrence in actual discourse.

5. The verbal setting with least effect on reduction of ambiguity is the one word preceding the word to be translated. The greatest effect is that of the entire sentence in which the word occurs.

6. A context consisting of one or two words on each side of the key word has an effectiveness not markedly different from that of the whole sentence.

7. The most important factors affecting contextual reduction of ambiguity are the accuracy

of the translators and whether the verbal setting includes words other than particles. The most practical context is therefore one word on each side, increased to two if one of the context words is a particle.

8. Under optimal conditions (most accurate translators, non-particle contexts, at least one word on each side of the key word) ambiguity is reduced to from 1/4 to 1/3 of the number of senses assigned to the word in isolation. A short verbal setting therefore reduces average ambiguity from about 5 1/2 senses to about 1 1/2 or 2.