Chapter Nine

THE SIGN-TEST FOR SIGNIFICANCE USED IN SUPPORT OF NEW CONCEPTS OF MEANING*

A scholar conducting research in the field of meaning faces the danger of accepting his intuition as a sufficient basis for drawing generalizations. Moreover, the theoretical framework he employs may influence his research to the degree that he regards as valid certain facts that hardly reflect linguistic reality.

To minimize the influence of these factors, empirical research coupled with statistical tests can be used to support or reject a given hypothesis. In the present paper I will demonstrate the extent to which the statistical test known as the 'sign test' can be relevant for semantic analysis.

As the analysis is based on a portion of an informant test designed and administered by C. S. Butler, I want to express my deep gratitude to him for both having agreed to my making use of the results of the test and being patient enough to introduce me to certain statistical concepts. Although this presentation is concerned with only a portion of the results of the informant test, I believe it necessary to start with a brief description of the project. Thus, it consists of 72 simple sentences of English. Each of the sentences refers to a person whose name is *John* and to whether he *is* or *is not at home*. The elements differentiating the sentences are the following: (a) the presence of a modal and/or a modal adverb and (b) the position of the modal adverb: initial, medial, and final. Ninety six subjects -- all native

---

* Kamila Turewicz, University of Łódź.

speakers of English -- were asked to listen to each of the sentences twice, decide upon the degree to which each to them believed *John is at home* and indicate it on the scale 0-100 by circling an appropriate value. The scores formed ordinal data.

For the purpose of this presentation I will concentrate on the results obtained for the following sentences:

(1) Possibly John´s at home.
(2) John´s possibly at home.
(3) John´s at home, possibly.
(4) Perhaps John´s at home.
(5) John´s perhaps at home.
(6) John´s at home, perhaps.

Sentences (1), (2) and (3) contain the modal adverb *possibly* and they differ only in the position of this adverb, i.e., initial in (1), medial in (2), and final in (3). In other words, as the sentences are formed of identical lexemes, there does not seem to exist any reason to anticipate differences in meaning among them. However, the analysis of raw data with the range of values between 0-9, as reflected in Table 1, for each of the utterances signals a lack of homogenity as regards the interpretations of the sentences.

No inference can be made until the data are sorted. The first step to organize the results is to arrange the values in frequency distribution tables, indicating the frequency of occurrence for each of the scores. These are next converted into a graphic representation such as a histogram or frequency polygon. Table 2 is the frequency distribution table for (1) and Fig. 1 presents the frequency polygon for the same sentence.

The frequency distribution tables uncover the existence of certain points around which the majority of the scores cluster, i.e., they indicate that the results tend to group around some values. For instance, for sentence (1) 61 out of 96 scores concentrate between the values 3-5, for sentence (2) 62 out of 95 scores have values between 3-5, but for (3) 60 out of 96 take values between 2 and 4. If we remember that the values reflect subjects´ interpretation of the utterances that differ only in the position of the adverb, we may at this point tentatively infer that due to the final position of *possibly* in (3) the assertive

Table 1

Raw scores for (1), (2) and (3); 96 subjects

N of S	(1)	(2)	(3)	N of S	(1)	(2)	(3)	N of S	(1)	(2)	(3)
1	6	6	5	33	3	1	4	65	1	1	4
2	6	5	7	34	5	4	3	66	5	8	4
3	4	3	3	35	3	3	2	67	4	5	4
4	4	6	2	36	5	5	5	68	9	4	3
5	5	5	6	37	1	4	5	69	6	5	4
6	5	5	3	38	5	5	4	70	2	4	3
7	3	2	2	39	5	4	3	71	5	4	2
8	3	2	3	40	5	6	5	72	5	5	3
9	1	2	1	41	4	5	4	73	1	2	4
10	4	4	5	42	4	3	3	74	5	2	5
11	2	3	3	43	4	2	1	75	4	4	4
12	6	6	7	44	4	3	3	76	4	3	5
13	3	4	5	45	4	3	3	77	3	5	3
14	5	5	6	46	5	4	4	78	2	2	2
15	4	6	5	47	5	5	8	79	5	5	4
16	4	5	7	48	4	5	2	80	4	1	2
17	3	5	2	49	3	0	1	81	1	1	0
18	6	6	4	50	3	3	2	82	5	5	1
19	5	5	4	51	6	3	3	83	4	5	3
20	5	4	2	52	5	5	3	84	2	4	1
21	5	5	5	53	5	4	3	85	4	3	2
22	3	2	5	54	5	5	7	86	5	6	5
23	2	2	2	55	1	3	5	87	2	3	5
24	2	2	2	56	3	4	2	88	8	5	2
25	2	2	3	57	4	6	2	89	6	7	6
26	4	5	2	58	6	4	6	90	6	5	1
27	4	6	4	59	3	2	4	91	2	3	2
28	0	10	1	60	3	4	7	92	2	3	1
29	6	7	7	61	2	4	4	93	5	-	8
30	5	5	4	62	1	3	2	94	6	6	4
31	4	3	3	63	7	8	2	95	5	5	3
32	5	3	2	64	2	1	3	96	- 2	5	1

Table 2

Frequency distribution table for (1)

Value	N of Sc.	Value	N of Sc.
0	1	6	11
1	7	7	1
2	13	8	1
3	13	9	1
4	21	10	0
5	27		

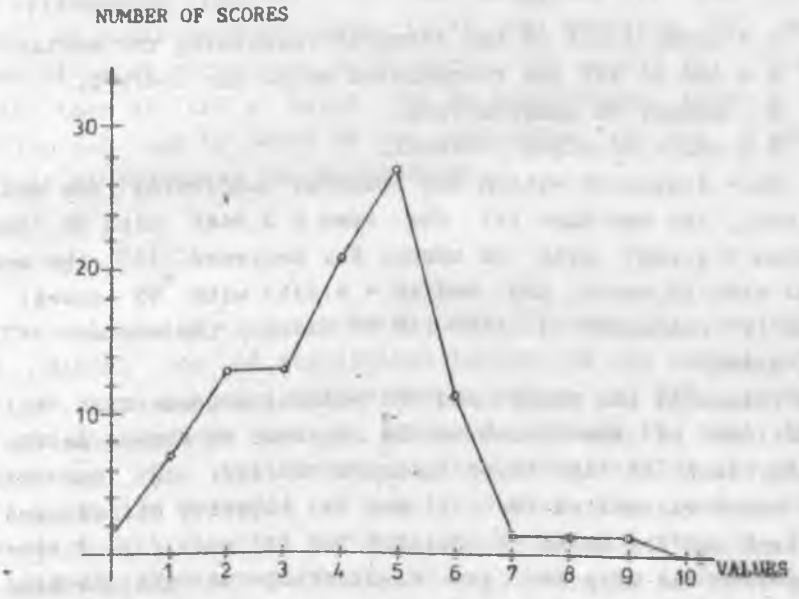

Fig. 1. Frequency-polygon for (1)

power of this sentence decreases. Yet, this does not suffice for formulating a hypothesis as regards the meaning of a sentence vs. the position of its adverb.

In order to obtain more precise information about the central tendencies in the distribution of the scores for each of the

sentences, we can employ the measures of central tendencies, namely, the *mean* and, as the data is ordinal, the *median*. To compute the means and the medians for (1), (2), (3) we use the expressions (a) and (b), respectively:

(a)

$$\text{mean} = \frac{\Sigma x}{N}$$

x = any of the numbers whose mean we are trying to compute,
Σx = sum of all the x's
N = number of observations.

(b)

$$\text{median} = L + \frac{N/2 - F}{fm} \; h$$

L = lower limit of the interval containing the median,
F = sum of all the frequencies below the limit L,
N = number of observations,
h = width of class interval,
fm = frequency within the interval containing the median.

Accordingly, for sentence (1) the mean = 3.9688 with 96 cases, the median = 4.1667 with 96 cases; for sentence (2) the mean = = 4.0842 with 95 cases, the median = 4.2353 with 95 cases; for sentence (3) the mean = 3.5218 with 96 cases, the median = 3.3095 with 96 cases.

The values of the means and the medians suggest that (a) the distributions of the scores can be regarded as skewed as the two values for each of the three sentences differ; (b) the medians and the means calculated for (1) and (2) slightly differ; and (c) the medians and the means calculated for (1) and (2) differ remarkably from the ones for (3). Considering on the one hand the fact that the taree sentences do not differ lexically and, on the other, the possible effect of irrelevant variables imbalance (subject variables, situational variables), we must not treat the obtained statistics for the samples as an exact reflection of the parameters for the populations from which the samples are taken. We can, however, employ a statistical test for significance to define the probability of obtaining the observed sample differences even if no difference between the values of the parameters for the populations can be established. In other words,

it is possible to find out whether the differences are significant for the populations.

The *sign test* is a convenient non-parametric test to be employed for our ordinal level scores whose distribution is skewed. As demonstrated in Table 3, while applying this statistical method we are concerned with the direction of the differences between the scores that are compared in order to obtain the number of pairs with the less frequent sign called x. Under the hypothesis that there is no difference between the parameters for the populations -- the null hypothesis -- we can anticipate that the number of pairs with positive differences between corresponding scores (marked '+') equals the number of those with negative differences (marked '-'). For sentences (1) and (2) the number of positive differences is 31 and negative, 35. As the number of pairs reflecting differences (N) equals 66, i.e., is greater than 25, the x value can be transformed into a normal distribution and by means of the expression (c) the z value for any pair of sentences can be computed.

(c)

$$z = \frac{N-2x-1}{\sqrt{N}}$$

The calculated test statistic for (1) and (2) has the value 0.37. Making use of statistical tables for the normal distribution we find out that this value points to a 35% probability that the differences can be observed although the null hypothesis is true. For the purpose of linguistic studies the difference is typically regarded as significant if the probability that it stems from irrelevant factors is not higher than 5% for both one tailed (directional in the sense that the investigator predicts the direction of the differences) and two-tailed (non-directional -- with no prediction as for the direction of the differences). The accepted probability level -- the level of significance 0.05 -- defines the critical value for z as 1.96. The test statistic falls below this value; therefore the null hypothesis cannot be rejected. In other words, the differences we obtained for the samples are not significant enough to regard them as reflecting similar differences in the populations.

Table 3

The sign test: the comparison of scores for (1) and (2)

(2)	(1)	sign	(2)	(1)	sign	(2)	(1)	sign	(2)	(1)	sign
6	6		2	2		3	0	+	1	2	-
6	5	+	4	5	-	3	3		5	2	+
4	3	+	4	6	-	6	3	+	4	4	
4	6	-	0	10	-	5	5		4	3	+
5	5		6	7	-	5	4	+	3	5	-
3	2	+	4	3	+	1	3	-	5	5	
5	5		5	5		5	5		2	2	
3	2	+	5	3	+	3	4	-	4	1	+
1	2	-	3	1	+	4	6	-	1	1	
4	4		5	4	+	6	4	+	5	5	
2	3	-	5	5		3	2	+	4	5	-
6	6		3	3		3	4	-	2	4	-
3	4	-	1	4	-	2	4	-	4	3	+
5	5		5	5		1	3	-	5	6	-
4	6	-	5	4	+	7	8	-	2	3	-
4	4	-	5	6	-	2	1	+	8	5	+
3	5	-	4	5	-	1	1		6	7	-
6	6		4	3	+	5	8	-	6	5	+
5	5		4	2	+	4	5	-	2	3	-
5	4	+	4	3	+	9	4	+	2	3	-
5	5		4	3	+	6	5	+	5		
3	2	+	5	4	+	2	4	-	6	6	
2	2		5	5		5	4	+	5	5	
2	2		4	5	-	5	5		2	5	-

In the situation when all three sentences are formed of the same lexemes it need not surprise anyone that no significant differences can be established among them. By the same token, regardless of the differences between medians for (1) and (2) on the one hand and (3) on the other, the null hypothesis can be formulated that there is no significant difference in the distribution of the parameters for the populations from which the samples are drawn as regards the interpretations of (2) vs (3) and

(1) vs (3). Accordingly, to verify the hypothesis the sign test statistics for (1) :(3) and (2) :(3) must be calculated.

For the former pair the number of differences with the less frequent sign equals 29; N (the number of pairs) equals 77; the computed z-score attains the value 2.05. Interestingly, this corresponds to 0.02 probability level, i.e., falls in the region above the critical value 1.96 associated with the 0.03 significance level for non-directional test. Also, the z-score for (2) and (3) attains a value higher than the critical one, hence the results of the test statistics are significant enough for rejecting the null hypothesis. Moreover, as the z-score for (2) and (3) equals 2.82, the corresponding probability level has the value 0.0025, which means that the differences under discussion are *highly* significant for both directional and non-directional tests.

Summing up, we can note the following: (a) the initial or medial position of the adverb *possibly* does not significantly influence the interpretations of the otherwise identical sentences; (b) the final position of the same adverb causes remarkable difference in the interpretation of this sentence vs the interpretations of those in which the adverb is *initial* or *medial*.

In the light of these facts it is tempting to generalize and put forward the argument that the same regularities can be established for other epistemic adverbs. Let us then use the same statistical methods with respect to sentences (4), (5) and (6). The median for (4) is 3.8, for sentence (5) it has the value 4.6944, and for (6) the median is 4.2272. The sign test results are even more interesting. Namely, the z-score for (4) and (5) equals 4.36, the value for the pair (4) :(6) is 2.09, whereas the one for (5) and (6) is 2.31. The analysis of the computed test statistics as well as the comparison of medians show that for *perhaps* the lowest assertive power obtains when the adverb is *initial* (it was the final position for *possibly*) and the difference in the distribution of the parameters for the populations appears to be significant in all three cases, though it is least significant for the *medial* and *final* positions.

The findings provide the basis for rejecting the suggested hypothesis that the position of an epistemic adverb itself is *regularly* responsible for the differences in interpretations of

otherwise identical sentences. Simultaneously, the results welcome the argument that sentences with the same lexical items differ significantly as regards their interpretations in relation to the order of the lexemes in the syntagmatic combination. Moreover, while comparing the values of the medians between sentences of the two sets we observe that the differences within certain pairs appear to be so subtle that we can predict the test statistics will point to no significant differences for these pairs.

Closing up this brief illustration of the extent to which the sign test technique can be useful in language studies I want to observe that it provides grounds for arguing against the concepts of meaning that assume a dichotomy between syntax and semantics. On the contrary, linguists should aim their studies at establishing, for example, what aspects of central meaning of an adverb are activated in relation to the adverb's position in a syntagmatic relation so as to result in the observed differences among the interpretations of these syntagmatic relations, remarkable enough to be perceived by native speakers of the language.