DOCUMENT RESUME

ED 124 181 IR 003 573

AUTHOR Pearson, Charls
TITLE Descriptive Summary of Georgia Tech's Semiotics Lab.
INSTITUTION Georgia Inst. of Tech., Atlanta. School of Information and Computer Science.
PUB DATE 76
NOTE 9p.; For a related document see IR 003 572

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Communication (Thought Transfer); Computer Oriented Programs; Computers; Higher Education; Information Processing; *Information Science; Information Theory; *Laboratories; Laboratory Equipment; Laboratory Procedures; Language; *Language Research; *Semiotics
IDENTIFIERS Georgia Institute of Technology; Natural Language; Research Laboratories

ABSTRACT

This document is a descriptive summary of the Georgia Institute of Technology's semiotics laboratory. A review of the goals and objectives of the laboratory is followed by a description of the facilities, including the computer software. The capabilities and uses of the laboratory are outlined for classroom experiments, instructional experiments, and research experiments. A review of the problems and needs of the laboratory concludes the document. (CH)

DESCRIPTIVE SUMMARY OF GEORGIA TECH'S SEMIOTICS LAB

by Charls Pearson

The following descriptive summary of the Georgia Tech Semiotics Lab was prepared for the ICS Ad Hoc Committee on Labs.

I. Goals and Objectives

A. Goals

The semiotics lab at Georgia Tech is currently designed to support a number of different courses of both general and technical nature. The experiments available in the lab have been chosen largely with these courses in mind, however several experiments are presently under way pruely for their intrinsic semiotic interest.

The lab is intended to not only educate the student with certain experimental techniques and empirical results in semiotics, but to instill in the student the classical attitude of other scientific labs, namely that of questioning nature. It is not sufficient merely for the instructor to tell the student that such and such a result holds, but the student should test the relation for himself. It is also intended that the development of such a lab will give us the opportunity to build up some sophisticated facilities with which it will be possible to carry out some semiotic experiments of a more scientific purpose.

B. Objectives

Student work in the semiotics lab is done for a number of purposes. Among the more important are the following:

1. To teach and to clarify the fundamental principles of semiotics. In this function the lab serves as an auxiliary to the lecture work of the course.
2. To stimulate the student's curiosity and to instill a scientific attitude; that of questioning nature rather than taking things for granted or accepting things on trust.
3. To emphasize the fact that a semiotic formula is a mathematical shorthand for expressing the results of an experiment.

4. To introduce the student to a number of measuring instruments and techniques, many of which are widely used in other fields such as linguistics, psychology, anthropology, sociology, and so forth.
5. To give practice in drawing conclusions and in thinking along scientific lines.
6. To give practice in writing a neat, clear report of experimental results.
7. To give practice in keeping a neat, clear scientific diary.

II. Facilities

A. General

Because of our school's computer, a heavy reliance is placed on computer techniques and instrumentation in most of the experiments. In contrast with more traditional labs, such as a physics or chemistry lab, which include a large physical facility for housing a number of students plus the equipment for running the experiments, the undergraduate semiotics lab has no such facilities. Since most of the experiments may be carried out on an individual basis and do not require the physical proximity of the lab instructor and the student, such facilities are not absolutely needed. The primary components of the lab are a collection of text samples and samples of other sign usages, and a collection of analysis programs and other computer-oriented semiotic instrumentation for the Burroughs B5700 Information System. There is a lab manual available which has a lengthly introduction in order to help orient the student who lacks a background of familiarity with scientific labs. This includes sections on the scientific diary or lab notebook, data and its treatment, calculations with significant figures, measuring instruments and techniques, and the lab report. Finally there is a small physical facility to house the graduate students who are working in the lab, their experiments, and their equipment.

Equipment in the lab consists primarily of several eidontic deviometers, several sets of iconic stimuli, and a borrowed projector and screen for exposing the stimuli and a borrowed timer for measuring response times. We anticipate being able to acquire a 2-field projection T-scope, an eye movement detector and an ontotic integrator in the very near future, for more refined experiments involving the perceptoin, interpretation, and memory of visual signs.

B. Hard Copy

For manual measurement and analysis

1.	Dictionaries	6
2.	Comparative Bibles (4 languages)	1
3.	Novels	2
4.	File Listings	4
	Totals	13

C. Machine Readable Copy

For automated measurement and analysis

		Corpuses	Reels
1.	American English	4	15
2.	Braille	1	1
3.	German	3	3
4.	Aphasic	1	1
5.	Phonemic	2	1
	TOTALS	11	21

D. Software (examples only, not exhaustive listing)

1. System Software
   a) File conversion routines for B5700
   b) File maintenance routines for the various corpuses
2. Mathematical Software
   c) Random number generators
   d) Chi-square and rank correlation analysis
   e) Non-linear regression and asymptotic estimation
3. Application Software
   f) Production of artificial word forms of orders 0-3 by random number generation
   g) Production of artificial word forms of all finite orders by random sampling of text.

h) Character count routines

i) all orders: monogram, digrams, trigrams, etc.

ii) all representations: graphemic, orthographic, phonemic, etc.

iii) all alphabets: Latin, German, Cyrillic, etc.

i) Rank-Frequency Analysis Routine with running type-token analysis from 1-50,000 tokens.

j) Production of artificial word forms of variable Markov orders.

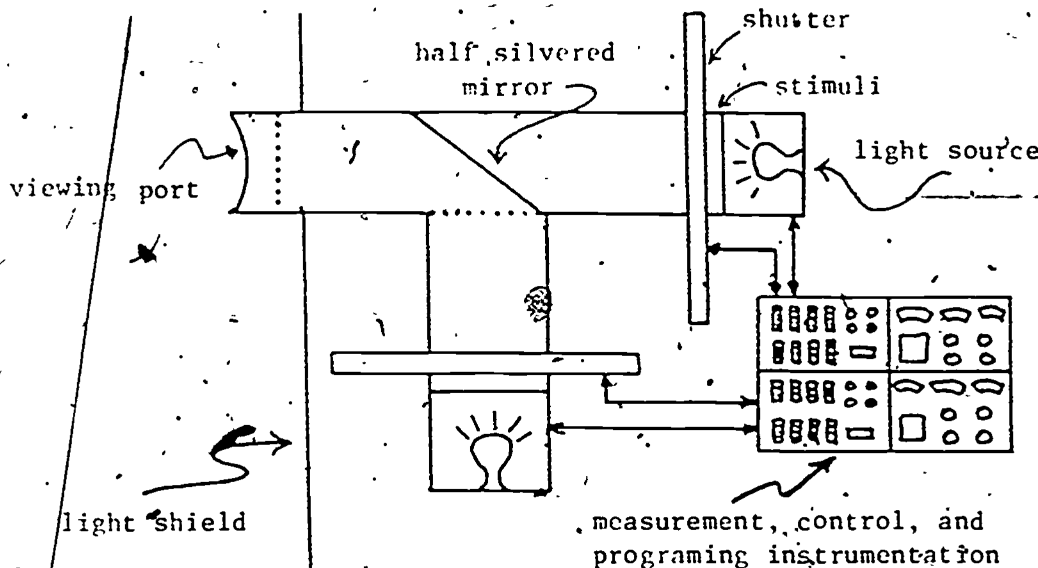
k) Articulatory features analysis.

E. Physical Facilities

1. One Lab Room ~ 15' x 15'
2. Two File Cabinets (for hard copy storage)
3. Four Desks.
4. One Tape Rack and 36 Reels of Tape
5. One Lab Work-Bench
6. One Response Timer (1 ms. to 1 min.)
7. One Carousel Slide Projector
8. One Tape Recorder
9. Three Eidontic Deviometers
10. One Projection Screen
11. Various Plug Boxes and Extension Cords
12. One (very rudimentary) Tachistoscope

F. Our Most Pressing Need is for a 2-Field Projection T-scope and Machine Readable Dictionaries (two on order), but we also anticipate the need of an eye motion detector and an ontotic integrator. Total funds required approximate $7,000.

2-Field T-Scope



III. Capabilities and Use

It is necessary to distinguish between experiments which have been developed, debugged, and now work smoothly on a mass production instructional basis; our conceptual capabilities for running instructional experiments; and our capabilities for doing original research.

A. Classroom Experiments

1. Law of Zipf and Estoup
   a) measurement of types and tokens in samples of natural language text
   b) number-frequency analysis
   c) rank-frequency analysis
   d) running type-token analysis
2. Language usage populations and inferential statistics
3. Information sources of Markov orders 0 thru 3 by random number generation.
4. Shannon's experiment to test Markov's Law, the increasing similarity of artificial words generated by random processes of increasing Markov orders to the shape of natural language words; $R = 0.88$, $\rho = 0.999$

5. The measurability of eidontic deviance.

B. Conceptual Capabilities for Instructional Experiments

1. All Zipf-Jung laws not involving meaning (including:
   a) Law of Zipf and Estoup
   b) Law of Abbreviation
   c) Fowler's Law of Repetition
   d) Jung's Law of Word Association
   e) But not Zipf's Law of Meanings or Baker's Law of Restriction.
2. Construction of evaluation of the eidontic deviometer.
3. Information sources of all finite Markov orders by random sampling of text.
4. Cluster analysis of word shape vs. mass sorting vs. dichotomous sorting.
5. Instrumental consistency problem for eidontic deviance.
6. Stability and significance testing of grapheme relative frequencies for various kinds of natural language text.
7. Statistical significance of various proposed equations for the Law of Estoup and Zipf.
8. Evaluation for various natural languages and for various styles of usage of ρ and θ in Mandelbrot's equation for the law of Estoup and Zipf:

$$P = C(r + \rho)^{-\theta}$$

9. Statistical significance of various proposed equations for the T/K function.
10. Use of the semantic differential.
11. Verification of Osgood's law of congruity.

C. Capabilities for Research Experiments

1. Evaluation of T/K ratio as function of age, language, psychological state, etc.
2. Evaluation of the length concept (i.e. determine the unit of lenght) in the law of abbreviation: length vs. frequency.
3. Evaluation of form of the length vs. number of meanings curve. Is this a law of types or tokens?

4. Herdan's contingency problem.
5. Isolation of the memory coding structure for cognitive memory. Presently trying to isolate the characteristics of iconic coding using iconic stimulus squares. When we are able to completely isolate these we will replicate the Bernbach experiment and measure the type memory accessed. The hypothesis is that it will be purely short term memory. Similarly symbolic coding —> long term memory and indexical coding and sensory memory. This opens up all of the recent experiments in cognitive psychology as tools for interpreting the semantic structure of signs.
6. Miller-Bruner-Postman Experiment and the measurement of Shannon's curve (the relative information per symbol vs. order of approximation to English). Requires both T-scope and Eidontic Deviometer.
   a) Shannon's curve: Relative Information per letter in English for letter sequences of various orders. So-called redundancy curve of English as a function of Markov order.
   b) M-B-P conclusion: Human information processing rate is a constant independent of relative shape or order of approximation.
   c) Procedure for measuring the Shannon curve:
      By first replicating M-B-P results with our improved equipment and verifying their conclusion, we can reverse their procedure and assume a constant information processing rate and measure the Shannon curve. Setting value of constant, Shannon gave upper and lower bounds. We can measure the curve with experimental tolerances much finer than Shannon's bounds.

IV. Problems and Needs

A. Accessing Sufficient Varieties of Text Samples

1. presently working with Dr. Rumbaugh of Georgia State in order to obtain machine readable corpus of non-human language use (Lana the Chimp).
2. negotiating a cooperative agreement with Georgia Mental Health Institute for machine readable samples of schizophrenic usage and other agencies for manic-depressive corpus. University of South Florida for additions to our machine readable aphasic file.

3. need help of colleagues for additional corpuses of different usages. Text book, medical records, etc.

B. Acquisition of equipment. To date acquisition of equipment has been mainly invent, build, beg, borrow, or steal. But funds are now required for purchase. Usually there are not instruments available for certain specific requirements and invention is a necessity. This requires extremely close and flexible cooperation with an instrument shop.

C. Space

To date the physical facilities are small and we would be crowded with one more graduate student. We must find a way to make more room.