Hillel Weiss
Bar Ilan University, Israel

Dynamic Indexing for the Analysis of Literary Text[1]

Weiss, H.: **Dynamic indexing for the analysis of literary text.** Int.Classif. 18(1991)No.4, p. 200-204
The article demonstrates theory and implementation of a dynamic indexing procedure for the analysis of a classic literary text (here the works of Sh.Y.Agnon). The three guiding principles for this procedure are: 1) *dynamic*, meaning the possibility to change data bases and improve them while reading, 2) *rapid access*, which is provided by multi-referencing channels leading to a specific point which is actually analyzed, 3) *maneuverability* of information into deductive and inductive channels, meaning *navigation* from text outside to a reference "inside" or from reference outside to the given text.
(Author)

1. Dynamic Indexing

The goal of the dynamic indexing of multi-strata texts is to control and identify all of the potential information and its significance that is absorbed in the text. This implies the creation of many flexible interlinked data bases (both small and/or large). These data bases can therefore be changed during the process of development.

The linking between the data bases is carried out by diachronic and/or analogic, hierarchic, or free access in order to attain the precise sense of the text. Since every encounter between the reader and the text can raise a different classification of data, the dynamic indexing and hence the index, must respond to the challenge of the complexity of the text.

One must in fact distinguish between verbal and non-verbal texts. Though our model is suitable for both, our interest lies in verbal texts and primarily in the most complicated ones such as classical literary texts. Examples of non-verbal texts are: The Human Body, Climate or a Military Organization. In short, any entity which is based upon some structure can be analyzed by this model. The more complicated the entity then the more useful the model will be.

According to the hermeneutic interpretation of text, a text is a system of relationships between a part, a unit and a total text. During the encounter with the text, units change their weight. Units which seemed marginal become central, and vice-versa.

Our model is aware of these problems and takes them into account.

Literary texts especially show a complex relationship between context and form. They demand from the reader both horizontal and vertical contemplation.

Up to the creation of this model, textual research was always limited although computer aided, as in the fields of lexicographic retrieval systems of hypertexts such as are known today. These systems cannot help in analyzing texts because their rigidity rules out any consideration of the dynamic character of the text.

The problem to be solved by the developers of this model was how to produce or to find software able to identify, index, classify, control, retrieve and analyze the vast information which can be gathered from a text.

When dealing with one file which is a database (that is, a classified output of the text), all the other data bases which have the relevant key words, synonyms, or abstractions of notions or subjects will receive the new data and reclassify them according to the new data being supplied.

In this system one can return and navigate to a specific place in the text and vice versa. One can move from a specific place in the text and link it to any channel of analysis that one chooses. When one decides to react to something in the text, a word, a phrase, or any signal which is found to have some meaning, one begins indexing and classing it according to channels, such as denotations and connotations; those can be all kinds of linguistic classifications or poetics or psycho-poetics and all kinds of other subjects.

Theoretically this vast information is not classified and collected by a single reader but by a whole culture. But the text must offer the opportunity to be linked to all kinds of classifications and must allow the reader to navigate within the information mass.

There are three conditions upon which the system is based:
1) Dynamic: meaning the possibility to change data bases, and improve them while reading.
2) Rapid Access: by multi-referencing channels to a specific point which is actually analyzed.
3) Maneuverability of information into deductive and inductive channels: meaning navigation from text outside to reference or from reference to the text.

The system combines the categorization by the tree sub-system, the hierarchic sub-system and the network

synchronic sub-system which is more associative and multi-directional.

This combination produces an "expert system" suitable to the purpose of classical text analysis.

2. The Process of Developing the Model

I spent a number of years in the field of information retrieval of literary Hebrew texts utilizing methods developed at Bar Ilan University. The essence of that methodology was to overcome the difficulties of Hebrew grammar and other linguistic problems encountered in data retrieval in Hebrew. Solutions were developed (in particular during Prof. Y. Choueka's term of office as head of the Institute of Data Retrieval with the aid of a team of researchers) by fulfilling specific requirements which satisfied the wishes of the literary researcher. These requirements included

1) the ability to retrieve information from within a single work among a totality of literary works;
2) retrieval of a textual unit such as a chapter or a verse;
3) organization of the retrieval of information in accordance with the publishing record of the work, and exclusive of its editing and entry into books or files.

The overriding purpose was to develop clearer concepts concerning the development of the work and other questions connected with identification of the relationships within the fabric of the text. One of the tools that were developed for analyzing texts is "family definition", which means editing an ad hoc thesaurus; the second tool is called "local feedback", which suggests "searchonyms" to the user that are related to the term of the query, once a set of relevant passages has been recognized.

In the course of applying these and other tools to texts, questions appeared and reappeared several times concerning the links between various searches, and the need as well as the technical ability to relate them to groups such as semantic fields of interest to the literary researcher. In addition, other questions reappeared such as the search for precise and broad meanings in the micro and macro text. The main problem which developed was how one can analyze and react spontaneously to the flow of a text.

As a result of these activities, a distinction was created between two kinds of principal users: first, the occasional browser who, like the user of a concordance, wishes to determine the location of a word or phrase or even a group of particular phrases; and second: the user who wishes to understand the entire text or to follow complex phenomena within or outside the text.

In 1988-89 we developed the concept of the need for a poetic, linguistic and analytic index to a text (1). This meant not only a collection of data produced and classified from within the text, but also a collection of data capable of serving purposes of textual analysis from the mere fact of its being a text given to decomposition, (or as it is known in the works of Dereida , deconstruction), classification, analysis and rechecking of conclusions according to the needs of various researchers: that is, treatment of, reference to and reorganization of classic multi-stratified text, dependent upon ancient literary tradition. Together with those qualities the model presents poetics and modern aesthetic concepts. Each of these textual characteristics requires specialized analytical tools and attention on the part of the researcher to the cultural-existential environment of the text.

The assumption is that the text contains an infinite number of units, bodies, etc., capable of objective identification (viz. objective and quasi-subjective) within it. Each segment or phrase in the text, each content and form can be linked to some kind of grammatical, syntactical, structural, historical, or psychological category, etc. and in parallel with a vast variety of approaches by each and every reader. The more the reader will recognize the textual data, the more firmly will his judgment be based. The resulting demand will always be unexpected, and therefore the tool for analyzing text must be capable of accepting identification and ascription of textual units to all kinds of classification categories, as well as their links to those in existence if the analyzer so desires.

Calling up data from within a text and its classification into groups is an elementary aid to the researcher. The simple procedure is to organize data bases according to topics such as geography, history, philosophy, etc. But if the researcher also has quick access to the citations as well as the ability to link them with other files of citations and to add notes and comments freely while differentiating between them and the text, we will have advanced another step in the computerized textual laboratory. If the citation files comprise not only citations of content but also poetic citations, which means that the question of "how" (form) is as important as the question of "what" (content), then we must maintain two main indexes, one of form and one of content, and create permanent links between them.

3. The Resulting Model

In the models shown below, we present a tool which permits an articulate reading of the text along with an immediate reaction to it, as well as a listing of each textual unit as defined in the index. They permit an ascription to each category or categories or idea or link to the external content of the text, the bibliography or to any need for a tool which a literary researcher might look for.. This includes all kinds of dictionaries. thesauri, dictionaries of synonyms in the text, bibliographies, biographical data synopses, etc.; in short, anything a reader can conjure up during his reading. These cumulative remarks are made both on the basis of categories set in advance, of a prepared model displayed herein, and also on the basis of a model which will be defined in the course of reading. The knowledge that it is possible to create a model, to retreat from it, to return to it in order to adjust it, etc. instills security and serenity in the

mind of the user of the index, since all of his reactions can be listed and tested in the light of previous findings, will not be lost, and can be easily recalled. Thus, an adequate solution to the question of navigating within a hypertext is presented here.

As an example of the above, we have chosen a very short piece composed of one phrase from the Work of S.Y.Agnon[2], (mentioned in article 1). With the aid of this piece we intend to demonstrate the difference between the model as a concept and idea (as described in article 1) as against the model in operation today utilizing the HIZE software (the Hebrew version of "IZE")[3]. This piece, which textually is highly compressed, denotative, and connotative, raises a host of poetic problems, questions of clarification of background for the emergence of the text.

Thus analysis of this text can become a starting point in the chain of the index of all of Agnon's works.

This article does not intend to deal with a particular classical writer but to demonstrate the model as a general system for treating complicated texts.

In the indexing of long text, we recommend that we proceed in three tracks. The first involves composition of a lexicon based upon retrieval of data of all concepts which seem popular or immanent to the work of a specific author: that is, whatever a textual researcher can conceive concerning an important subject within the context of the entire writings of a given author (clearly consulting with other outstanding authors). The second track is the expansion of a lexicon by searching for conspicuous concepts from within the culture in which the author created his work. All of these are searches from within the macro of the text. The third track, in parallel, focuses upon the strict work on the micro of the text: the super close reading which must also be circumscribed within considerations of efficiency: that is whatever is desired by the average researcher.

Notes

1 This research was made possible by the assistance of the Israeli National Academy of Sciences, the Foundation for Basic Research.

2 Shmuel Yosef Agnon, a Nobel Prize Winner in literature 1966, is considered to be the major Hebrew writer (1887-1970) whose works reflect generations of Hebrew-Jewish-European culture by webbing all the sources into a hyper-modern, post-Kafka creation. One of its main components is the complexity of genders. This monumental creation has stimulated generations of critics, researchers, and interpreters to try to decode his works. Thirteen years ago, while decoding a long complex text of Agnon's, composed of hundreds of citations and quasi-citations, I realized that without a computer it would be impossible to control the poetics and their ingredients and structure.
Agnon lived in Germany during 1913-1924. The first edition of his four great volumes appeared in Berlin in 1931. Some of his major novels deal with his experiences during that period. Several of Agnon's books were translated into the German language.

3 The operative application of the theory and requirements of the system were implemented on HIZE software with the aid of Tzvi Ilani and the assistance of and improvements carried out by Avraham Yosef.

Reference

(1) Weiss, Hillel: An analytical index for the complete works of Agnon: A scholarly tool in preparation. Literary & Linguistic Comput. 4(1989)No. 3 p.169-173.

Prof. Hillel Weiss, Bar Ilan University, Ramat Gan, Israel.

Appendix

Screens 1 to 26 as they were displayed during the process, translated from the original Hebrew text.

```
-----------------------------------------------------------------
                         screen 1
           DYNAMIC INDEXES  TO S.Y.AGNON*  >

                  A scientific tool for data
               management in a multi-level text

                 Retrieval and Analysis (1991)

NOTE 1: > implies calling up next screen

* all the works of Agnon consist of 2,500,000 words stored in
  a mainframe fully retrieved.

-----------------------------------------------------------------
                          screen 2
< DYNAMIC INDEXES

A.  List of texts and dates of publications  >

B.  General bibliography for works of Agnon  >

C.  Indexes for all Agnon's works   >

D.  Indexes for a single text >>

NOTE: using the > moves you to the following screen
-----------------------------------------------------------------

-----------------------------------------------------------------
                          screen 3
Hereby Represented. Choose One of The Stories: TEARS >>
-----------------------------------------------------------------

-----------------------------------------------------------------
                         screen 4

        1. SUBJECT INDEX >

        2. POETIC INDEX  >

        3. NAMES & PLACE INDEX >

        4. BIBLIOGRAPHY INDEX >

NOTE: THE INVERTED CARROT (<) RETURNS YOU TO THE PREVIOUS SCREEN
```

screen 5

< POETIC INDEX

1. GENRES >>
2. MOTIFS >
3. SOURCES >> (Refer to screen #6)
4. TITLES >
5. OPENINGS >
6. ENDINGS >
7. DEFINITIONS >
8. THE NARRATOR AND HIS FUNCTIONS >> (Refer to screen F)
9. CLUES >
10. CHARACTERISTICS >

screen 6

Poetic Index < Sources

Bible >

Wisdom Literature >

Exegesis >

Mysticism and Hassidism >

Poetry >

Books of Discourse and Instruction >

Questions and Answers >

Unidentified Sources >

screen 7

(Entered from screen #5)
< GENRES

1. FOLKTALES >
2. FABLES >
3. PROVERBS >
4. ANECDOTES >
5. NOVELS >
6. SHORT STORIES >
7. MEMOIRES >>

screen 8

< SUBJECT INDEX

1. HISTORY >
2. GEOGRAPH >
3. ECONOMIC >
4. LANGUAGE >>
5. PSYCHOLOY >
6. LAW >
7. PHILOSOPHY >
8. RELIGION >
9. SOCIETY >
10. LIFE CYCLE (BIRTH, DEATH, MATURITY ETC.) >> (Refer to screen F).

screen 9

< LANGUAGE

1. COLOCATIONS >
2. ABBREVIATIONS >
3. STYLISTIC MARKINGS >
4. SYNONYMS & THESAURUS >

EXAMPLE: SHORT ANALYSIS OF A TEXT BY THE NAMED INDEXES

screen 10

"TEARS" >> (Refer to screen 13)

It was told to me < by Rabbi Samuel Aryeh.< When I was young< I dwelt< in the village of Kishlevitz.< This is the village where the Baal Shem Tov< (before he was revealed)< was a ritual slaughterer.< I found there an old ritual slaughterer<, over eighty years old<. I said to him, "maybe you know someone who once knew the Baal Shem Tov?<" He told me, " A Jew who saw the Baal Shem Tov I didn't find. A gentile who saw the Baal Shem Tov, I found<. When I was young I lived in a certain gentile farm<. Every time I poured water on the sharpening stone< he would shake his head<.
I thought that he did this because of old age<. One time I felt that he did this as an insult<. I asked him. "Why are you shaking your head while I am working?" He replied. "You don't do your work properly<. Yisraelki< when he sharpened his knife used to wet the stone with his own tears<."

Note: Each carrot (<) is linked to another specific screen or screens that are automaticly editted by the software.

screen 11

Context of the Single Word - OCCURANCES OF THE WORD "TEARS" IN THE CONTEXT OF THE COMPLETE WORKS OF AGNON >

screen 12

A List of Hebrew Occurences - The Word "Tears"

screen 13

POETIC INDEX> NARRATOR> RUMOUR

> BIBLIOGRAPHY > WEISS > BETWEEN * STYLE * AND MEANING

IT WAS TOLD TO ME > (Comnes from screen #10)

WEISS. 47/48 PRECIS: SEARCHES FOR "TOLD ME": THE DIFFERENCES BETWEEN THESE EXPRESSIONS APPEARANCE IN THE TEXT CONTEXT OF THE SINGLE WORD -COLLECTION OF APPEARANCES IN AGNON'S WORKS >> (Back to screen #10)

screen 14

AN EXPANOED GROUP OF "TEARS": A POETIC INOEX (OTHERS)
(total 29)

< Motifs
- a. motifs index poetic tears
  - 1. Testimony > (2)
  - 2. Oral Testimony >
  - 3. (others) > (2)
- b. revelation >
  - 1. A righteous man * publicized > (3)
- c. Testimony* oral >
- d. Sharpening >
- e. other >

2. Narrator >
- a. Tears > poetic > index > narrator >
  - 1. tradition >
- b. (other) (2) >

3. Genres (4) >

4. Headlines (3) >

5. Others (2) >

The Following is an Example for identifying and linking Biblical sources to the text.
Hebrew and in English:

---

screen 15

Ezekiel >> (Goes to screen #16)

Job >> (Goes to screen #18)

Ecclesiastes >> (Goes to screen #25)

---

screen 16

Ezekiel XVI / 14 > And thy renown went forth among the nation for thy beauty:

---

screen 17

Link

A phrase from the text:
whose.name.was.known.all.over.the.world

(textually: its.renown.went.forth.all.over.the.world)

Ezekiel XVI/14 And thy.renown.went.forth.among.the.nations.

for thy beauty:

Back to the text >>

---

screen 18

(Entered from screen 15)

Job XXIX/4 > as I was in.the.days.of.my.youth.when.GOD shielded my tent:

---

screen 19

Contents of "Poetic Index of Sources (total 25)
I. Index Tears Poetic Sources >
A. Our Sages >
1. Talmud >
a. according to >
(I) tractate Naziree >
b. Others (2) >

2. Midrash Commentaries (Homiletical Exegesis) >
a. Chapters * according to Rabbah * Eliezer Heger Horev >
b. Midrash * Tanhuma Buber >
c. others (2) >
B. Bible >
1. Hagiograph (4) >
2. Sources Index Poetic Tears >
a. Bible (3) >
b. Job >
3. According to * Prophets (2) >
C. Exegesis (2) >

---

screen 20

Contents of "Talmud" (Total of 5)

1. Index >Tears >Poetics >Sources >Sages >Talmud
A. According to
1. Naziree >
B. Naziree
C. Tractate Blessings (Berakhoth) >

---

screen 21

(Entered from screen 14)
< Motifs > Testimony >

Old * Man's * Testimony
Link

Gentile

Back to the text - first >

second >

---

screen 22

Link

When.I.was.young (textually: In.the Days.of.My. Youth)

Job XXIX/4 : as I was in.the.days.of.my.youth, when GOD shielded my tent:

Back to the text >>

a paradisiac experience !

---

screen 23

(Entered from screen #7)
Subject index > Life Cycle >

Childhood >

Old . Age >

Youth >>

---

screen 24

Subject index > Life Cycle >
Poetic index > Memory >

When . I. was . young > (textually: in. the. Days . of. My. Youth)

---

screen 25

(Refered from 15)

Ecclesiastes VII/28 > One man among a thousand I have found; but a woman among all those I have not found.

---

screen 26

Link

A Jew who saw the Baal Shem Tov I.did.not.find., a gentile who saw the Baal Shem Tov I.found.

Ecclesiastes VII/28 One man among a thousand I.have.found; but a woman among all those I.have.not.found.

Back to the text > (Screen #10)
