

# AN INTRODUCTION TO EXPLICATIONS DE TEXTE - BY COMPUTER - OF GOLDEN OLDIE LOVE SONGS

by Marvin Green

Daniel Boorstin, in a 1961 book, *The Image*, tells of a mother who has just received a compliment on her baby's beauty. "Oh, that's nothin'," the mother replies, "You should see his photograph."

Text is a representation of human thought, and it is quite different from the thought it represents. Thought exists in the head. Text exists in so-called "alphanumeric symbols," meaning marks of punctuation, spaces, letters and numbers.

We are in the computer age, and a very important question is whether a machine can explicate text.

To answer, you need to investigate the exact difference between the underlying thought and the text which expresses it. When you explicate thought, you are answering the question, What does a person mean? When you explicate text, the question is, What do the alphanumeric symbols mean?

Consider the statement: "A woman's a two-face,/A worrisome thing who'll leave ya t' sing the blues in the night." Johnny Mercer wrote that statement. It is an assertion about women.

If you want to know Johnny Mercer's thoughts on women, you need to know things about Johnny Mercer. For example, he was married just once. But you can't really tell from that bit of information what he personally thought about women.

When you limit yourself to Johnny Mercer's assertion as it is expressed in text, you find that it appears, of course, in his song *BLUES IN THE NIGHT*, and it is preceded there by additional text. The various alphanumeric symbols in the whole section are the following:

My mama done tol' me when I was in knee-pants, My mama  
done tol' me, Son!  
A woman'll sweet talk, and give ya the big eye, but when the  
sweet talkin's done,

A woman's a two-face,  
A worrisome thing who'll leave ya t' sing the blues in the  
night.

Without a machine, you can explicate the symbols in that larger context. You can see, for example, that the symbols impute the assertion about women to a first-person narrator who may or may not actually exist (i.e., "I") and who may or may not be Johnny Mercer himself. You can see that this narrator in turn attributes the assertion to another person who may or may not actually exist (i.e., "My mama"). Also, you can see that the symbols provide a time dimension by the device that the narrator is speaking right now (i.e., is speaking the narration itself) and is telling about something (i.e., what "My mama done tol' me") which previously (i.e., "when I was in knee-pants") he may or may not actually have experienced.

When you read further into the alphanumeric symbols of *BLUES IN THE NIGHT*, you find that the narrator repeats the "woman's a two-face" assertion word for word, but this time the assertion is the narrator's own assertion. He is making it contemporaneously as a part of his narration:

From Natchez to Mobile, from Memphis to St. Joe, wherever the four  
winds blow;  
I been in some big towns an' heard me some big talk, but there is one  
thing I know,  
A woman's a two-face,  
A worrisome thing who'll leave ya t' sing the blues in the night.

And at the end of the song, you find the narrator explicitly connecting his own assertion with his present memory of his mama's assertion. He says, "My mama was right, there's blues in the night."

Well, the once-married Johnny Mercer died in 1976. Perhaps during his lifetime he had been in Natchez, Mobile, Memphis and St. Joe, or perhaps not. He was never divorced. He had two children. He had a mother. Did he think a woman's a two-face? As Sam Goldwyn says, "The answer is a definite maybe."

And as for the text you are considering, which is *BLUES IN THE NIGHT* itself, it is a work of fiction that either is or is not autobiographical. The point is that, if you

read and understand and appreciate BLUES IN THE NIGHT, you do not really know anything at all about Johnny Mercer or about his mother, and you do not really have to know anything about them.

Or maybe you do. The columnist Ann Landers seems to believe that you can tell what an author thinks from what the author says. Someone asked her to track down the source of a certain remark: "Men have much happier lives than women. They marry later and die earlier." She replied, "The man who wrote that was H.L. Mencken." Then she added, "Incidentally, very little is known about his married life. His statement suggests it wasn't anything to brag about."

You ignore Ann Landers at your own risk. Although her comment on Mencken was gratuitous, she makes a living by dispensing practical wisdom. So here, too, in this inquiry into machine explication of text, she reminds you of the obvious: what an author writes must be directly connected - somehow - with what the author thinks.

No less a person than Aristotle agrees with Ann Landers. He explains that Johnny Mercer had a universal mental experience which he could put into spoken words, and then he could put the spoken words into alphanumeric symbols. Here is how Aristotle says it:

Spoken words are the symbols of mental experience and written words are the symbols of spoken words. Just as all men have not the same writing, so all men have not the same speech sounds, but the mental experiences, which these directly symbolize, are the same for all, as also are those things of which our experiences are the images. . .

So text involves a conventional part, the alphanumeric text symbols, and a universal part, the underlying experience. Some conventional alphanumeric text symbols are elemental (such as "b," "l," "u," "e" and "s"). Some are speakable words and phrases (such as "a whooee duh whooee," which also appears in BLUES IN THE NIGHT). These speakable words and phrases are so-called "strings," meaning sequences of consecutive elemental symbols, including spaces and marks of punctuation. While particular elemental and string symbols are conventional for a particular language, that which they symbolize is universal.

So it was Johnny Mercer who wrote the string "a woman's a two-face" for English, but everybody - specifically including Johnny Mercer, Ann Landers,

Aristotle, Sam Goldwyn, H.L. Mencken, Tolstoy, Dante, Mao Tse-tung - more or less knows what it is like to have a person "sweet talk, and give ya the big eye, but. . . leave ya t' sing the blues in the night."

Here's how a machine - a computer - fits into this text picture. A computer is a natural for "reading" text, because a computer can recognize and differentiate the "b" and the "l" and the "u" and the other conventional elemental alphanumeric symbols.

Using this ability, you can, for example, cause a computer to ascertain the number of words in a text. Your computer can "read" one elemental symbol at a time and can count the ones which stand for spaces and can apply the formula (NUMBER OF WORDS = NUMBER OF SPACES + 1). Thereby your computer can determine that "A woman's a two-face" consists of four words. Furthermore, noting the apostrophe in "woman's," your computer can determine that that string really is a contraction for the two words "woman is." Therefore your computer can ascertain that the four words "A woman's a two-face" may be regarded as five words which say, "A woman is a two-face."

But Ann Landers and Aristotle claim it is not sufficient to manipulate the conventional symbols. You have to deal with the underlying universal. Specifically, they are arguing that your computer has to have experience of its own. So they are protesting about your computer what the envious traveling salesman protested concerning Professor Harold Hill in *THE MUSIC MAN*, namely, "But he doesn't know the territory!"

Their reasoning is that text is more or less like frozen concentrated orange juice. A processor removes water, pulp, seeds and volatile esters and aldehydes from natural orange juice. A consumer tries to reconstitute the original by adding water. When Johnny Mercer processes his mental experience into alphanumeric symbols, you as a consumer try to reconstitute it into your own mental experience. You do so by adding your own knowledge about the specific conventional symbols – about "two-face," for instance, and your knowledge derives from your own experience with "two-face." If you do not have any experience, or if you are relying on a computer to do the reading, you necessarily will lose something from that which Johnny Mercer was thinking. The February, 1987, Consumer Reports supports Ann Landers and Aristotle in this contention:

Despite forty years of trying, the orange-juice industry has yet to come up with a processed juice that tastes as if it came fresh from an orange. Something is always lost in the translation. That something is primarily the zesty, eye-opening, sweet taste of a fresh orange, a stimulation of tongue and nose easy to discern but hard to describe.

Now you have to determine whether Ann Landers and Aristotle and Consumer Reports are right or wrong, and to do so you can investigate the subject of love as it is processed into text.

First of all, there is a clean dichotomy of machines and persons with respect to love. Whereas the world is a long way from having a computer which has experienced love, virtually every human being has had substantial experience. Second, there is abundant material for you to use. Perhaps half the world's text involves love.

You can investigate using any text at all involving love, but you will find sufficient useful material in Golden Oldies - popular songs of the middle 1900s.

Most popular songs are about love. After all, love is an emotion such that bursting into song is common and easy to do. And from the standpoint of the writer trying to make a living, love songs have the merit that there is a universal market for them. Johnny Mercer published more than 1000 songs, Cole Porter about 1000 and Irving Berlin over 1500, maybe 3000. Although Irving Berlin frequently wrote on topics that are not specifically about love - for example, show business, reveille, Easter bonnets, and Christmas, also he wrote about love in connection with such things as America, guns and champagne.

Indeed it is difficult to keep love out of popular songs. Thus, patriotism is explicit in Berlin's GOD BLESS AMERICA, and patriotism is a species of love ("land that I love"). Likewise, in Mack Gordon's CHATTANOOGA CHOO CHOO, which essentially suggests an overnight journey from inside Pennsylvania Station in New York City to outside a railroad station at Chattanooga, Tennessee, the song ends with its narrator's making a statement that presumably involves love:

Shovel all the coal in, gotta keep it rollin', Woo woo,  
Chattanooga, there you are.

There's gonna be a certain party at the station, Satin and lace, I used

to call "funny face."  
She's gonna cry until I tell her that I'll never roam, So Chattanooga Choo  
Choo won't you choo choo me home?  
Chattanooga Choo Choo won't you choo choo me home?

Another thing commending popular songs for your investi- gation is that their text is brief - usually 100 words, more or less, seldom as many as 200. BLUES IN THE NIGHT is quite long. It has 192 words. BEGIN THE BEGUINE has 206. THREE LITTLE FISHES is a rarity. It has 335.

Whereas a novel about love will have tens of thousands of words and numerous twists and turns, a song, because of its brevity, must be pretty much to the point. That is, you will necessarily find in each love song a usually coherent statement on some one love topic. For example, you could say, as a first approximation, that BLUES IN THE NIGHT, when you boil it down, makes the single statement, Watch out for love! there's blues in the night. This coherent statement feature might facilitate your investigation.

One possible problem with using popular love songs is that the various statements they make may be of dubious intellectual quality or validity. What with love a universal experience, you have to concede that anybody and everybody could with equal justification claim authority on the subject. Your machine could very well find itself wandering through labyrinths of error.

So what are Johnny Mercer's special credentials for warning, "Watch out for love! there's blues in the night?"

And what about Aristotle? Well, some 2500 years ago, Aristotle studied under Plato, and Plato devoted his SYMPOSIUM to love. Even though Plato is still esteemed as a philosopher, does he have any better love credentials than the rest of us? It may be that he does. He was a mouthpiece for Socrates. And Ann Landers will tell you that Socrates and Mrs. Socrates, whose name was Xanthippe, make H.L. Mencken and Mrs. Mencken seem the happiest of love mates.

Anyway, in more recent times, the likes of Sigmund Freud and Ann Landers herself and Dorothy Parker also have written authoritatively about love. Of these, Dorothy Parker is the most candid pretender. She said,

Oh, life is a glorious cycle of song, A medley of extemporanea.

And love is a thing that can never go wrong, And I am Marie  
of Roumania.

So, granting that Everyman can write or sing love songs, there is an easy solution to the problem that the songs might not be intellectually sound: you don't actually care. Maybe a woman is a two-face; maybe a woman is not a two-face. Since all you want is to know whether a computer can explicate the alphanumeric statement "A woman's a two-face," neither underlying validity nor underlying invalidity makes any difference.

How wonderful it would be to demonstrate that a computer can explicate the 15,000 or so words which Plato authoritatively wrote in the SYMPOSIUM. Your goal is modest. It is simply to demonstrate that a computer can cope with the 100-word statement of any typical Golden Oldie or the not-quite-200-word statement of BLUES IN THE NIGHT or any other statement of any other length that appears in any other Golden Oldie.

So here's the plan, and it is really simple. You ask your computer questions about the song. If it can answer just about any reasonable question accurately and relevantly, you will have to acknowledge that it understands the song in some way and to some degree.

For example, you will want to ask your computer, Does the song have a first-person narrator or a third-person? This question is important. Whereas a first-person narrator himself participates in the action or circumstance of a song, as is the case in BLUES IN THE NIGHT, a third-person narrator claims no such participation. Thus, in AS TIME GOES BY, the narrator is not a participating character at all. Perhaps you can characterize him as a philosopher, because he makes remarks such as,

You must remember this, a kiss is still a kiss, A sigh is just a sigh;  
The fundamental things apply, As time goes by.

It is easy for your computer to discriminate between first-person and third-person narrators. In each song it simply counts the number of strings that constitute the first-person pronouns "I," "me," "my," "mine," "myself," "we," "us," "our," "ours," "ourselves" and "ourselves." Here are representative results

from selected songs:

<u>Song</u>	<u>Number of first-person strings</u>
IN THE MOOD	22
BLUES IN THE NIGHT	14
SMOKE GETS IN YOUR EYES	12
STORMY WEATHER	11
BEGIN THE BEGUINE	11
THREE LITTLE FISHES	2
AS TIME GOES BY	1
THE SWINGIN' SHEPHERD BLUES	0

So, when you ask your computer, Does THREE LITTLE FISHES have a first-person narrator? it replies, No. When you ask about BEGIN THE BEGUINE, it says, Yes.

All you do is adopt a reasonable formula. You instruct your computer to give the "yes" answer whenever there are more than two first-person strings in a song and to give the "no" answer in the other cases.

If, following your instructions, your computer occasionally makes a mistake, you can be quite forgiving. After all, to err is human.

Anyway, specifically to exemplify the whole project, consider BLUES IN THE NIGHT. It is, or was, a popular love song. It is comprised of 1,021 elemental alphanumeric symbols clustered into 192 speakable words. You are going to use a computer that doesn't know the territory, that doesn't know what love is or "knee-pants" or "mama." You are going to determine whether this inexperienced computer can tell you what the 1,021 alphanumeric symbols are saying.

Appendix A sets forth a full statement of the protocol for the entire project. Appendix B gives detail concerning the programming methods and discusses problems and refinements.