AN ANALYSIS OF TOPIC INFORMATION IN PARALLEL WRITTEN AND SPOKEN EXPOSITORY TEXTS IN ENGLISH

Gudrun E. Sherman
Webster University, St. Louis

The majority of linguistic studies that compare spoken and written language tend to examine syntactic and lexical features or to investigate contextual influences on spoken and written language and the relationship between the two modes of language (see Akinnaso 1982, Chafe and Tannen 1987, Macauley 1990). This paper reports on a study that is interested in determining characteristics of information distribution and development, rather than syntactic properties, in spoken and written texts that share similar contextual factors. Of particular interest in this analysis is the portion of information in a sentence that represents what the sentence is about, given its context. This unit is called topic information here, and it is to be distinguished from units of information that supply information about the topic and are often referred to as comment in other literature (e.g. Hockett 1958).1

In this analysis, topic information is the kind of information that tends to recur in consecutive sentence units in the paragraphs of written texts as well as the "paratones" of spoken texts (Brown 1977:86). If the topic information of a given sentence is not reiterated in a subsequent sentence, then the information unit functioning as the topic information in that following sentence signals a shift in topic focus (Faigley 1986). In one of the written texts, for example, Governor Reagan was introduced as new topic information in the text, signalling a shift from one viewpoint to another in a paragraph on the various viewpoints of individuals that participated in the debate on the decriminalization of marijuana in California in the seventies.

Although it tends to be introduced at the beginning of sentences (e.g. William 1989), topic information, as it is defined here, can occur in other sentence positions as well. Compare the topic information, which is underlined, in the following three sentences from one of the written texts:

2. Decriminalization in Maine, California, and Ohio was preceded by extensive legislative staff research.

4. In Ohio and Maine, decriminalization was part of an overall criminal code revision, which ... .

81. In only three states which eventually decriminalized marijuana was there partisan conflict.

Why investigate topic information? What insights can be gained from such an analysis?

(1) As the unit that functions as the topic information in a sentence tends to recur as topic information in following sentences, the topic information units can be compared in each text. A comparison of topic information units may shed light on how topic information is developed in
different written and different spoken texts and how this development is a
reflection of the contextual influences of the texts' discourse topics and/or
generic features.

(2) As topic information represents what authors consider a suitable
common ground about which something else can be said, the information that is
chosen as topic information can provide clues about writers' or speakers' assumptions and expectations about their audiences' background knowledge. Thus, an analysis of topic information may provide insights about a correlation between sentential information distribution and a text's contextual features, such as its discourse topic, purpose, and target audience.

(3) As speakers are expected to speak fluently and continuously, without
disruptive pauses and hesitation, the selection of topic information becomes a
crucial point in maintaining information flow in speech (Ong 1982:40). Given
this observation, an analysis of topic information may indicate how certain patterns of information distribution and development accommodate the pragmatic and cognitive constraints of the medium in which a text is produced or received.

Given these three considerations, an analysis of topic information in
different written and spoken texts may prove to be a promising approach in
identifying characteristics of information distribution and development that
reflect discourse-specific factors such as discourse topic, purpose, and
audience as well as medium-related constraints such as (1) the lack of
immediate feedback from the audience in writing, (2) the evanescence of the
spoken word in speech, and (3) limited time for thinking in speech (Gumperz et
al. 1984, Ong 1982, Chafe and Danielewicz 1987). In order to be able to
distinguish between discourse-related differences and medium-related
differences between written and spoken texts, this analysis uses written and
spoken texts that are controlled for contextual discourse factors (see
Macaulay 1990). Categories of information from the works of Halliday (1967,
1985), Chafe (1987), and Lovejoy and Lance (1991) are used to describe the
various ways topic information is distributed and developed in the sentence
units of these written and spoken texts.

A. Corpus of Data

In this analysis, two pairs of parallel written and spoken texts are
examined that are expository in nature. Each pair of written and spoken texts
shares similar content, fulfills the same general communicative purpose, and
is targeted at similar or the same audiences. The first pair of texts deals
with the decriminalization of marijuana in 13 U.S. states; they represent
summaries of facts and factors related to the decriminalization processes in
these states; and they are both addressed at professional audiences that share
an interest in the subject matter. The second pair of written and spoken texts
deals with the measurement of crime in the U.S; they provide explanations of
various methods of crime measurement and evaluations of their usefulness, and
they are intended for student audiences that are enrolled in an introductory
sociology course. All four texts are produced or co-authored by the same
professor of sociology.
B. Method

In the two written texts, the basic sentence unit for the analysis is the T-unit as defined by Hunt (1964). A terminable sentence unit in the spoken texts is determined on the basis of falling intonation, a pause, and a sense of finality, not necessarily syntactic boundaries (see Chafe and Danielewicz 1987:95).

The topic information in the sentences is analyzed with respect to two questions:

1. How does it function within the sentence?
2. How does it function within the discourse?

In order to answer the first question, one must determine how the topic information is treated in a sentence. Is it treated by the writer or speaker as given information or as new information? If treated as given, topic information typically occurs at the beginning of intonation units as above in sentence (2) and (4). If treated as new, topic information is typically placed towards the end of intonation units where it receives unmarked informational focus (Halliday 1967, 1985). For an example, see sentence (81) above. In some cases, the entire topic information unit is treated as given, yet individual lexical elements within the unit function as new because (1) they may introduce new ideas to the discourse and/or (2) they are marked with informational focus.

In order to answer the second question, one must determine how the topic information relates to the rest of the text. Does the topic information that is treated as a given in a sentence represent an active or a semi-active concept? (Chafe 1987:31). If given topic information has been previously mentioned, then the information is considered given\active. If topic information treated as given has not been mentioned before in the text, yet can be inferred from the context of the text, then the information functions as given\semi-active.

Does information that is treated as new within topic information represent an inactive concept (Chafe 1987:31), or is it cohesively related to preceding information? (Halliday and Hasan 1976:4). If information functioning as new has not been mentioned before and introduces a new concept to the discourse, then the information is considered new\inactive. If information functioning as new within topic information has been mentioned before or is semantically related to previous information, then the information is marked as new\cohesive (Halliday 1985). If the entire topic information unit is treated as new, then the topic information simply functions as new.

Here are examples of the various functions of topic information in the texts:

1. Given\Active Topic Information (G-A)

2. Decriminalization in Maine, California, and Ohio /// was preceded by extensive legislative staff research. (G-A)
4. In Ohio and Maine // decriminalization was part of an overall criminal code revision which diverted attention from the marijuana penalty changes.

(2) Given\Semi-Active Topic Information (G-S)

41. Later articles // dealt with a statement by the director of the White House Special Action Office for Drug Abuse Prevention.

42. He /// advocated decriminalization of personal use of marijuana.

G-S--------G-S

43. But Governor Reagan /// warned against easing marijuana penalties in an address to the California Council on Criminal Justice.

(3) New\Cohesive Elements within Topic Information (N-C)

9. Now, apparently it’s generally been assumed /// that all people in the area should be included in these calculations.

N-C G-S-------------G-S G-A

10. Every man, woman, and child in the area /// should be included in this total population figure.

(4) New\Inactive Elements within Topic Information (N-I)

1. Earlier we discussed the arbitrary nature of what is termed “crime.”

N-I G-A

2. The definition of crime varies not only across countries but across states as well.

(5) Topic Information Treated as New (N)

37. Keith Stroup pointed out /// that “there are FBI statistics that show 420,700 persons were arrested on marijuana charges in 1973.”

The data on the various functions of topic information in the written and spoken texts are presented in Table 1 below. In the analysis of the data, the chi-square test is used to determine the statistical probabilities for the differences in number to occur by chance. Because the four texts vary in length, the total numbers of items in each text are converted into numbers per 100 sentence units. These are the numbers to which the chi-square test is applied. They appear in boldface type in the table. The written and spoken texts on the decriminalization of marijuana are abbreviated as W:DM and S:DM, and the written and spoken texts on the measurement of crime are abbreviated as W:MC and S:MC.
C. Analysis

Table 1

The Functions of Topic Information in the Written and Spoken Texts

<table>
<thead>
<tr>
<th></th>
<th>W:DM</th>
<th>S:DM</th>
<th>W:NC</th>
<th>S:NC</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-A</td>
<td>69/72</td>
<td>48/51</td>
<td>94/84</td>
<td>98/79</td>
</tr>
<tr>
<td>C-S</td>
<td>29/30</td>
<td>22/23</td>
<td>24/21</td>
<td>20/16</td>
</tr>
<tr>
<td>N-C</td>
<td>22/23</td>
<td>8/ 8</td>
<td>43/38</td>
<td>47/38</td>
</tr>
<tr>
<td>N-I</td>
<td>64/67</td>
<td>16/17</td>
<td>60/54</td>
<td>27/22</td>
</tr>
<tr>
<td>N</td>
<td>9/ 9</td>
<td>22/23</td>
<td>5/ 4</td>
<td>26/21</td>
</tr>
<tr>
<td>TOTAL</td>
<td>96</td>
<td>95</td>
<td>112</td>
<td>124</td>
</tr>
</tbody>
</table>

Similarities between the Parallel Texts

According to the table, most units that function as topic information consist of given/active information, i.e. previously activated information. There is relatively little difference between the parallel written and spoken texts in the numbers of units that function as given/active information (p < 0.100 and p < 0.750). There is also little variation between the parallel texts with respect to their use of given/semi-active topic information (p < 0.500 and p < 0.500). These findings support the common definition of topic as a structure representing given information (e.g. Halliday 1967). Whether a text is written or spoken, the topic information in each text tends to consist of given information. The only difference in this study is that in the written texts, given topic information tends to be expanded with new information, especially new information that has not been mentioned before (see below).

The topic information that represents given/semi-active information tends to refer to similar subject matters in the parallel spoken and written texts. Many of the given/semi-active topic information in the written and spoken texts on decriminalization represent organizations and individuals who actively participated in the debates about the decriminalization of marijuana (e.g. Jerry Brown, Reagan, Chief of Police Davis, law enforcement, Republicans, Communists).

The fact that these terms are used as topic information, not as part of comment information, suggests that the writer and speaker expects his audience to be familiar with these organizations and political figures. He treats them as "points of departure" rather than introducing them as new information in the comment (Halliday 1967:200).

In the parallel written and spoken texts on crime measurement, man, woman/women, child/children, infants, mayor/mayors function as inferrable topic information in passages that further explain and illustrate previously
given definitions of crime measurement or show the flaws of these measurements by means of concrete situations. Because of their common use in everyday communication, these topics tend to be more accessible to student audiences than some of the topic information in the other pair of texts would be (e.g. NORAHL, the Assembly, they as a general reference to the legislators in a particular state).

The parallel written and spoken texts show similar numbers of given\semi-active topic information units, which reflect similarities in the kinds of information chosen as topic information. But the texts also contain given\semi-active topic information that is not found in the other parallel written or spoken text. Given\semi-active topic information such as I, you, and we occur only in the two spoken texts, enhancing the speaker's "involvement" with his subject matter and his audience (Chafe and Danielewicz 1987:105).

Given\semi-active topic information that is found only in the written text on decriminalization refers to facts and factors that influenced the decriminalization processes in the various states (e.g. the press, the boundary between major and minor possession, the charges, a review, medical opinion). In addition to the terms of everyday usage mentioned above, the written text on crime measurement also contains topic information that can be inferred from the general context or "the scenario" of crime (Sanford and Garrod 1981:114): evidence, act, judges, courtroom, correctional institutions, officials. The spoken text on crime measurement also uses similar terms related to the context of crime, such as defender, witnesses, facts of the case, but they are not used as topic information.

Medium-Related Differences between the Parallel Texts

The two medium-related differences between the spoken and written texts of this study seem to be that (1) in the written texts, given topic information tends to be expanded with new information, especially new\inactive information, and (2) the spoken texts contain more units of topic information treated as new information than their parallel written texts (p < 0.025 and p < 0.001). Both written texts use considerably more elements of new\inactive information in their topic information than their parallel spoken texts (p < 0.001 and p < 0.001), and the written text on decriminalization uses more elements of new\cohesive information to expand given topic information than its parallel spoken text (p < 0.010).

By expanding given topic information with new information in the two written texts, the writer introduces new information in the topic unit as well as the comment, which is the unmarked unit for new information. Thus, chunks of new information may appear at least twice in a written sentence unit in a fairly short interval. Since it takes time to select a new idea that is appropriate to the context of what has been said before, the topic information in the two spoken texts is less likely to be expanded with new information than the topic information in the two written texts.

What kinds of topic information are expanded with new information in the two written texts?

In both written texts, the new information that is used to expand given
topic information introduces new information about the subjects that are of
main interest in the texts: decriminalization bills, laws, penalties, as well
as states and drug users in the written text on decriminalization; crime or
crimes (acts); measurement (denominators, numbers, increases, decreases); the
reporting of crime, communities and areas, or violators of the law in the
written text on crime measurement.

In what ways is new/inactive information integrated into the topic
information in the two written texts?

(1) the head noun of a compound subject (e.g. the SEVERITY of existing
penalties, a brief review of the legislative HISTORY of the
decriminalization bill; the OCCURRENCE of a specific crime,
ANOTHER CONSEQUENCE of prosperity)
(2) the periphrastic phrase of a compound subject (e.g. MAYORS of many
American METROPOLITAN AREAS, the problem of MISSING INFORMATION),
(3) the prenominal modifier of a subject (e.g. SUCCESSFUL
decriminalization bills, the DEMOGRAPHIC, CULTURAL, AND POLITICAL
characteristics of these states; A SEPARATE reporting of all
arrests, THIS APPARENTLY CONTRADICTORY message of both increases
and decreases, EVEN the REPEATED, RELATIVELY VIOLENT acts of
middle- and upper-class youths), and
(4) the postnominal modifier of a subject (e.g. The effort to
LIBERALIZE the state's marijuana law; The evidence REQUIRED to
PROVE the same type of crime)

In the spoken text on decriminalization, new information is introduced
with subjects that are similar to the ones in the written text, but not as
frequently as in the written text. The subjects that are expanded with new
information are also states, decriminalization, and people involved in the
debate (e.g. NONE of these states, THE ROLE of these law enforcement people,
people who SUPPORTED decriminalization, ONE of the OTHER things that is
INTERESTING about this decriminalization business). However, the information
functioning as new has often been mentioned before rather than representing
new/inactive information. If new/inactive information is integrated into the
topic information, it is usually integrated as a postnominal modifier, not a
prenominal modifier (e.g. people who SUPPORTED decriminalization, one of the
other things that is INTERESTING ...).

In the spoken text on crime measurement, content similar to the one in
the written text is expressed in three different ways:

(1) If new information is integrated into topic information, the new
information is marked with audible stress (e.g. the OCCURRENCE of
a specific crime; ONE ARGUMENT, on the other hand, FOR using court
records). The stress on the new information may enable the
listening audience to process the unit of topic information piece
by piece instead of having to process the whole unit.
(2) Instead of integrating the new information into the topic, the
speaker expresses the new information as a comment about the
topic (e.g. "the increases were down to 7% per year" versus the
written version of "an 8 PERCENT increase in crime in a given year
actually represents a reduction ...", or
"In fact, some police departments, because of these pressures,
don't report the truth about the magnitude of crime in their community..." versus the written version of "The problem with official police records is that the TEMPTATION to ALTER figures is very great").

Instead of beginning sentences with complex topic information, the speaker introduces the topic information towards the end of intonation units where it receives phonological prominence through unmarked information focus (e.g. "it may not be a good idea / to TRY to CALCULATE CRIME RATES at all" versus the written version of "The PROBLEM with these alternative denominators is that they do not fit most Americans' concept of crime, ...)."

This strategy of treating topic information as new information by placing it towards the end of intonation and sentence units appears to be a more common strategy in the spoken texts of this study. It accomplishes the dispensing of information into two separate intonation units so that one idea per unit can be presented to the listeners (Chafe 1987:31). The corresponding comment precedes the topic, often in an intonation unit of its own, so that both the comment, may not be a good idea, and the topic, to try to calculate crime rates, may receive phonological prominence. The following examples from the two spoken texts illustrate the partitioning of topic and comment information into two successive units instead of integrating them into one:

21. So it doesn't make any SENSE ... / to INCLUDE the TOTAL POPULATION of ALL MEN, WOMEN, and CHILDREN in the COMMUNITY / in calculating say rates of armed robbery).

An Alternative Information Distribution:
The INCLUSION of the TOTAL POPULATION of ALL MEN, WOMEN, and CHILDREN in the community doesn't make any SENSE in calculating say rates of armed robbery.

29. ... / it's a BETTER IDEA / to USE the NUMBER of PARKING PLACES / because, as some of you know, ... .

An Alternative Information Distribution:
Using the number of parking places is a BETTER IDEA.

13. ... / it was really an UPHILL BATTLE / to PASS DECRIMINALIZATION LAWS / because ... .

An Alternative Information Distribution:
To PASS decriminalization laws was really an UPHILL BATTLE.

54. In New York / it was also a LONG, DRAWN OUT PROCESS / to GET the LAW CHANGED.

An Alternative Information Distribution:
In New York / getting the law changed was also a LONG, DRAWN OUT PROCESS.

The above examples show that the topic information that is treated as new information tends to consist of phrases that are more complex than simple noun phrases. If used in an alternative information distribution, the topic would
have to be complemented by the comment in the same intonation unit. That is why treating complex topic information as new information would be better suited to speech, given the pragmatic constraints of speech that determine an optimal information load in a unit, i.e. one new idea per intonation unit (Chafe 1987:31).

**Discourse-Related Differences between Texts of the Same Mode**

In addition to these two medium-related differences between the parallel texts, the table also shows large variations in number between the two spoken texts in their use of given\active topic information (p < 0.025) and their use of new\cohesive information in topic information (p < 0.001). The first difference may be a reflection of the dyadic form of the spoken text on decriminalization, a message-centered conversation, which fosters the use of semi-active topic information such as I, you, and we instead of previously mentioned topic information referring to conceptual subject matter, such as crime. The second difference reflects discourse-specific features of the spoken text on crime measurement, which it shares with its parallel written text.

The two spoken texts differ in the way their discourse topics are developed. The spoken text on crime measurement is characterized by a topic development that is conveyed by the frequent recurrence of key words such as crime, calculating, crime rates, police, and report. When given topic information is expanded with new information in the written and spoken texts on crime measurement, the information is often related to the measurement and calculations of crime (e.g. NO burglary, MOST countries, MANY types of crime, ALL people, EVERY man, woman, and child; ONE person, TWO murders, 100,000 people, TOTAL population), or the information has been mentioned before in connection with a different method of crime measurement (e.g. 35. The PROBLEM with these alternative denominators is ... 61. ONE PROBLEM with official police records is ...).

The pervasiveness of crime measurement throughout these two parallel texts may explain the higher numbers of new\cohesive elements of information in the topic information of the written and spoken text on crime measurement.

It should be mentioned, however, that the difference between the two written texts is statistically much smaller (p < 0.100) than the difference between the two spoken texts (p < 0.001).

**D. Conclusion**

Although the findings of this study are limited to a relatively small sample and although they cannot be used to make predictions about other written and spoken expository texts, the texts of this study provide sufficient data to suggest that a similar analysis of topic information could prove useful in determining medium-related and discourse-related differences between spoken and written texts. To strengthen the validity of the medium-related differences identified in this study, more studies are necessary that examine topic information in parallel written and spoken texts on different subjects and of different genres such as narrative, description, exhortation, instruction. (see Macaulay, 1990).

The results of this analysis confirm findings and observations from
previous studies. For example, the tendency that new information is integrated into topic information in the written texts supports qualities such as "integrated" or "synoptic," which have been attributed to written language by Chafe (1982:50) and Halliday (1987:75). The fact that similar information is expressed in the comment in the spoken text and integrated into the topic in the written text supplies further evidence to Halliday's (1987:78) observation that in spoken texts information tends to be rendered as a process and in written texts as a product or an objectified process.

However, unlike previous structural analyses, this analysis focuses on the description of topic information. It does not consider, for example, the information realized by nominalizations, attributives, or prepositional phrases that are used as part of comment information. Unlike previous syntactic analyses, this analysis is able to identify meaningful differences among the written and spoken texts in their use of attributives by determining their informational functions within the discourse (new or inactive or new cohesive) and their placement within the topic information unit (prenominal or postnominal). Thus, this analysis of topic information contributes new insights to the study of spoken and written language as well as supports previously made observations and suggestions.

NOTES

1 Topic information in this analysis is to be differentiated from the sentence-initial syntactic category of topic. It is comparable to the concept of topic entity, a term used and discussed in Brown and Yule (1983:137-8).

REFERENCES


Galliher, J.F. and Murphy, S.E. 1986. The process of marijuana
Hunt, K.W. 1964. Differences in grammatical structures written at three grade levels. Tallahassee: Florida State University.