情 報 メ デ ィ ア 15－1
（１９９４．　３．　１１）

話題構造認識を用いた映像検索システム

竹下 敦
NTT ヒューマンインタフェース研究所
take@aether.ntt.jp

情報に対して人間が認識する話題構造に着目し、話題構造を用いて目次と索引を提供する映像検索システムを提案する。これらの機能により、映像情報の概要把握と情報の必要な部分へのアクセスが可能となる。さらに、対話、モノローグ、テキストという３種類の伝達形態の言語データから話題構造を認識する手法を提案する。本手法は、「人間は言語コミュニケーションにおいて意味的情報を伝達にするためには話題構造を伝達する必要があり、そのためには何らかの手掛かりを用いたり、言語現象を伴う」という仮定に基づき、伝達形態に応じた手掛かりや言語現象を規則化して認識を行なう。また、評価実験によって本手法の有効性を示す。

A Topic-oriented Video Retrieval System

TAKESHITA, Atsushi
NTT Human Interface Laboratories
1-2356 Take Yokosuka-Shi, Kanagawa 238-03, Japan

A topic-oriented video retrieval system is proposed. It provides a table of contents and indexes created from topic structure. Users can grasp an outline of the video library, and can play back video sequences connected with an interesting topic. This paper also proposes an approach to topic structure recognition that is based on the hypothesis that linguistic clues and observable phenomena are used for communicating topic structures, and they can be used to develop recognition rules. The validity of the approach is shown by comparing manual and system topics.

1 はじめに

情報のマルチメディア化によって、テキスト、音声、映像などのメディアを必要に応じて組み合わせ、より適切な形で情報を表現することが可能となってきた。また、それに伴い、より効果的な情報交換やヒューマンコミュニケーション(以下、HC と呼ぶ)への期待が高まっている。

しかしながら、マルチメディア情報はそのまま保存しただけでは、その情報を再利用することは困難であり、使い捨てになってしまう。情報を蓄積しておいて、それを有効に再利用するためには、情報に対してある種の構造付けを行なっておくことが必要である。

本稿では、そのような構造として、人間が情報に対して認識する話題構造に着目し、話題構造を利用した映像検索システムと、そこで用いる話題構造認識処理について述べる。

2 話題構造と情報検索

2.1 言語情報の基盤としての話題構造

人間にテキストや対話データを与えて、「同じことが書いてあるブロックと、その『同じこと』を求める」という課題を与えると、個人差なく同じ構造を答えるという性質が実験的に確認されている[竹下 93]。図 2-1 はこのような構造の例であり、これを「話題構造」と呼ぶ。話題構造は入れ子構造を形成するので、話題を示す「話題語」とそれがどの文からどの文まで継続するかという「話題スコープ」によって表現できる。

これまで、自然言語言語理解や合成の研究分野では意図構造[Grosz86]や議論構造[Cohen87]、修辞関係理論[Mann90]、主題構造[Danes74]などの構造が提案されている。しかしながら、これらの構造が顕著であるような言語データは限定されるので、本研究では多くの言語データに対して明確である話題構造に着目する。

話題構造は HC において言語情報の意味を限定するために用いられるものと考えられる。例えば、「ポスト」という語は、「郵便」の話題の中で用いられれば「郵便ポスト」を指すが、「企業内のリストラ」の話題の中で用いられれば「地位」を指す。また、他義語に限らず、一般の言葉の意味も話題構造によって、明確になるものと考えられる。

これは人間の記憶方式による。人間は知識を整理した形で記憶しているわけではないので、知識を用いる際には、まず記憶の中から関連のある知識を集めるために「話題構造」を用いて焦点化を行い、次に集めた知識を必要な形に再構成する。

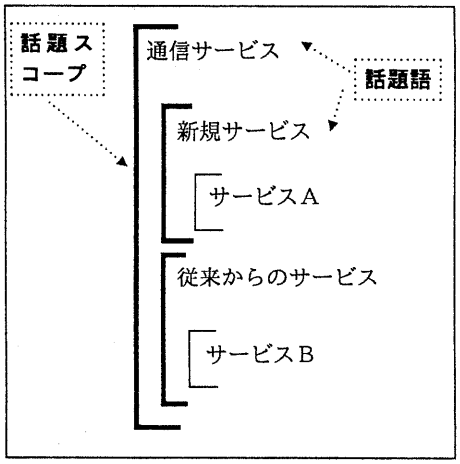

図 2-1　人間による話題構造認識の例

2.2 話題構造を利用した映像検索

我々は「言語情報付き映像データ」に対して 3 章で述べる話題構造認識処理を行ない、その結果得られた話題構造を用いて、目次と索引を提供する映像データ検索システムを SUN の SS/2 上に試作した。図 2-2 に示すように、入力としては、文字化された対話データと、発話番交代と映像のフレーム番号との対応表を SUN 上に、映像データを LD 上にあらかじめ与える。

話題構造認識モジュールは与えられた文字化対話データから話題構造を認識する。目次・索引インタフェースは話題構造を目次としてX-window上に表示し、さらに話題構造からの索引作成とその表示も行なう。索引は索引語だけでなく、前後の文章も付加したKWIC(KeyWord in Context)と呼ばれる形式を用いた。KWICの例を表2-1に示す。

目次・索引インタフェースで、ある話題語や索引語がユーザに選択されると、LD制御モジュールは、話題スコープ情報を話題構造から抽出し、「発話番-映像フレーム番号対応表」を用いてそれを映像情報のフレーム番号に変換し、その区間の映像だけを再生する。

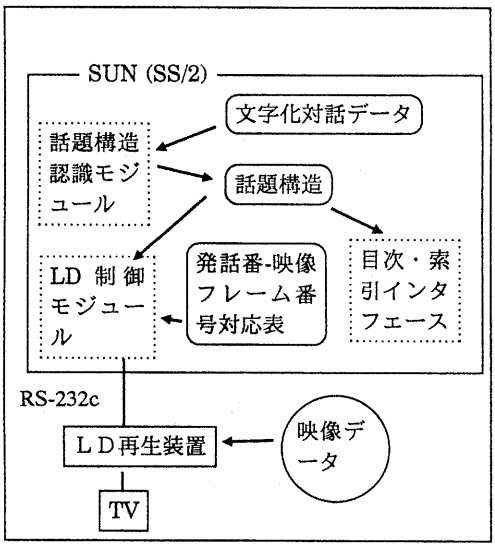

図2-2　映像検索システムの構成

前文章	索引語	後ろ文章
ット。日本一巨大な	エンジン	は、２１世紀
。この世界一小さな	エンジン	は、温度差に
一、いや世界一小さい	エンジン	なんですよ。
野さん。まず。はい。	最初	の日本一は。
、見て下さい。これ。	紙	の切れ端みた
……	……	……

表2-1　KWIC索引の例

話題構造を用いることによって、人間にとっての情報圧縮ができるということが、本システムからの知見として得られた。すなわち、話題の目次により、人間は大まかな話の流れを把握することができる。また、話題スコープ情報によって、各話題がどこから始まって、どこまで継続しているかということが分かるので、必要な情報にだけアクセスが可能であるという意味でも、情報圧縮が可能である。

このシステムが仮定している「言語情報付き映像データ」は日本ではまだ、あまり流通していない。しかし、米国ではクローズド・キャプションと呼ばれる、文字情報であるキャプション信号を映像信号に多重化して字幕を表示する放送方式が普及しており、キャプション付きの番組が非常に多い[坂本91]。日本でも、難聴者対策として今後、普及すると考えられるので、「言語情報付き映像データ」は入手しやすくなる。また、現在でも、文字起こし作業が従来と比較してコストが下がっているので、そのようなデータは比較的安価に作成することは可能である。

本システムは、映像データ以外にも利用できる。本システムが行なっていることは、映像データのように構造を持たない「ベタ」の情報に対して、目次と索引を与えて「本」と同じ構造を持たせることによって、人間にとって利用しやすいようにすることである。したがって、例えば、雑誌の取材のためのインタビューや、会議や講演会は、最近では文字起こしされることが多いので、有望な対象である。

3 話題構造認識へのアプローチ

3.1 話題構造認識の着眼点

2.1で述べたように、話題構造はHCの基底である。したがって、話題構造を相手にうまく伝達するために、話し手や書き手は何らかの言語的手掛かりを意図的に用いたり、あるいは結果的に言語現象が生じるはずである。例えば、対話では話題が大きく変わるときの手掛かりとしては「次に」などの手掛かり句を

用いたり、新しく導入した話題を明確にするための言語現象として、質問-応答等のやりとりが行われることがある。本研究では、このような手掛かりや言語現象に着目し、どのような手掛かりがどのような目的に用いられるかを規則化し、それにより話題構造認識を行なう。

3.2 一般化話題構造認識モデル

まず、話題の入れ子関係を無視して考えると、話題展開は図3-1に示すようになる。すなわち、ある話題が提示されて、確立される「話題確立区間 GS(Grounding Section)」と、その話題が維持される「話題維持区間」の繰り返しである。対話の GS では、例えば質問-応答、確認-応答等のやりとり系列が見られる。また、話題の大きな切れ目には「次に」などの手掛かり句が用いられる。

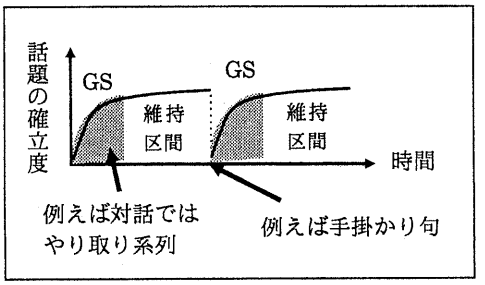

図 3-1　話題展開過程での確立区間と維持区間

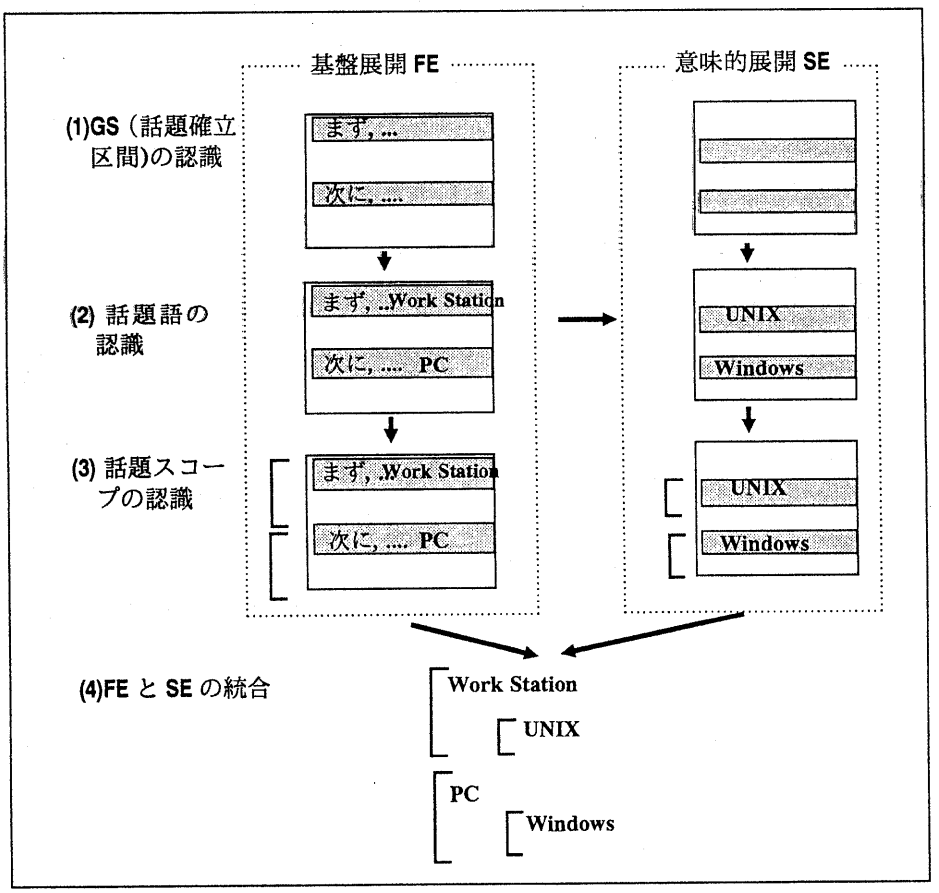

図 3-2　一般化話題構造認識モデル G-TREC の概要

次に、話題の入れ子関係も考慮して、話題展開を考える。大局的な話題構造を伝達するためには、「初めに」「次に」のような手掛かり句が用いられるが、このように明示的に示された話題構造を「基盤展開 FE」(Fundamental Expansion)と呼ぶことにする。

ところが、全ての話題構造が手掛かり句によって示されるわけではない。これは、手掛かり句の入れ子が深くなり過ぎると、手掛かり句の対応を取るのが困難になり、かえって聞き手や読み手の負担になるからである。したがって、基盤展開の中では明示的手掛かりなしに、より小さな話題が展開されるが、これを「意味的展開 SE」(Semantic Expansion)と呼ぶことにする。

図 3-2 に話題構造認識モデル G-TREC (Generalized Topic Structure RECognizer) の概要を示す。まず、基盤的展開 FE について、GS, 話題語、話題スコープの認識を行ない、次に、意味的展開 SE について同じことを行なう。そして、最後に 2 つを統合する。なお、話題スコープの認識は、話題の始点と終点だけでなく、話題間の入れ子関係も認識することである。というのは、話題スコープの入れ子関係が、そのまま話題の入れ子関係となるからである。

3.3 ヒューマンコミュニケーション伝達形態の分類

我々は表 3-1 に示すに示すように、話題展開に影響を与えると思われる記録性とインタラクションの有無によって、HC 伝達形態の分類を行なった。ここで、モノローグとはテレビニュースや講演会のように 1 人で話す形態である。本研究では、対話、モノローグ、テキストを対象とする。

		インタラクション	
		○	×
記録性	○	e-mail 等	テキスト
	×	対話	モノローグ

表 3-1 HC 伝達形態の分類

表 3-1 によると、対話とテキストは共通した性質を持たないが、モノローグは対話ともテキストとも共通した性質を持つことになる。これは、歴史的に考えて、対話で基本であったのが、その後、モノローグ、テキストと発展したという流れにも合致する。さらに、表 3-2 に示すように、話題展開様式に関しても、モノローグは他の 2 者の中間的性質を持つ。なお、表 3-2 については、3.4 で説明する。

対話	モノローグ	テキスト
論理構造なし		論理構造あり
インタラクションあり	まれに自問自答	
逐次型話題提示	一括型話題提示	

表 3-2 伝達形態による話題展開様式の違いの例

3.4 話題構造認識処理の例

HC 伝達形態によって異なる言語現象が起こることは報告されているが[Oviatt89]、その原因まで体系的に追及した研究は少ない。本研究では、話題展開という観点から、そこで用いられる手掛かりや言語現象を検討する。

3.4.1 基盤展開 FE の認識例

FE における GS の手掛かりとして、対話とモノローグでは表 3-3 に示すような手掛かり句が用いられる。これに対して、表 3-2 で示したように、テキストでは言語情報の記録が可能となったので、章立てや章タイトル、箇条書き等の論理構造が話題展開の道具として利用可能となった。FE-GS はこれらの手掛かりを探すことによって認識する。

種類	手掛かり句の例
入れ子開始型	まず、第 1 に、最初に、...
話題転換型	次に、ところで、第 2 に、...
入れ子終了型	最後に、終わりに、...

表 3-3 手掛かり句の例

また、FE-GS での話題提示は、対話では複雑な名詞句を使うことができないので、表 3-2 に示したように、話題を少しずつ提示する「逐次型」が用いられる。これに対し、テキストでは、その記録性によって複雑な名詞句を使うことが可能となったので、「一括型」が用いられる。モノローグでは両者が用いられる。

したがって、モノローグの場合は逐次型と一括型のどちらであるかを同定する必要がある。話題語は図 3-3 に示したような顕著名詞句マーカによって提示されるが、例えば、FE-GS の最初の文に複数の明示マーカが含まれていれば、逐次型とみなす。

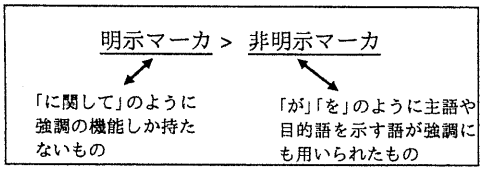

図 3-3　顕著名詞句マーカの優先順位

次に、FE-GS における話題語を認識する。各 GS において、図 3-3 で示した明示マーカで提示された名詞句か、非明示マーカによって示されているが固有名詞を含んだ名詞句を話題語として選ぶ。ここで、固有名詞を優先するのは、固有名詞は指示する対象が具体的に決まっているので、発話内容が指す事柄を限定する能力が強く、話題の機能に適しているからである。

もし、話題語の候補が複数検出された場合は、対話とモノローグでは、時間的に最も早く発話された名詞句を話題語とする。というのは、発話内容を正しく伝達するために、FE 話題はできるだけ早く提示されていると考えられるからである。これに対して、テキストにおいては複雑な名詞句を用いることが出来るので、他の名詞句を連体修飾していないもので、時間的に最も早く提示されたものを話題語とする。

次に、話題スコープを話題レベルとともに認識する。最も外側の話題のレベルを 1 とし、入れ子の 1 つ内側に行くにつれてレベルが 1 ずつ増えるものとする。対話とモノローグの場合は、表 3-4 に示した規則によって、手がかり句の種類に応じて話題レベルを決定する。話題スコープの始点は、FE-GS の始まりとし、終点はそのレベル以下の話題が開始する直前とする。

今回の手掛かり句

前回の手掛かり句	開始	転換	終了
開始	+1	0	0
転換	+1	0	0
終了	+1	-1	-1

略語の意味
開始: 入れ子開始
転換: 話題転換
終了: 入れ子終了

図 3-4　基盤展開でのレベル付け規則

3.4.2　意味的展開 SE

対話における SE-GS は、質問-応答等のやりとり系列、同じ名詞句の繰り返しなどで示される。これは[Shegloff79]において提案されている“summons-answering sequence”と呼ばれる会話開始・話題導入機構に合致する。

モノローグやテキストでも、新たな話題を確立するためには、ある程度の期間が必要であると考えられる。テキストの場合は、長い段落の先頭を SE-GS 開始点とする。これに対して、モノローグでは段落がない。しかし、例えば「これは」などの話題継続句によって文単位のマージを行ない、疑似段落を求めることによって、テキストと同様の方法で GS 認識を行なうことが出来る。

SE-GS における話題語の認識方法は、対話については FE-GS に対するものと同じである。これに対して、モノローグとテキストに関しては、1 つだけ相違点がある。その相違点とは、明示マーカで示されたり固有名詞を含んだりする名詞句だけでなく、直前の FE-GS に含まれている名詞句も優先するということである。これは、モノローグとテキストにおいては、話し手や書き手が話題展開権を保持できるので、相手に分かりやすく伝達するた

めに、これから話したり書いたりすることを、FE-GS において要約した形で予告することができるからである。

話題レベルに関しては、モノローグとテキストでは、SE の話題レベルは 1 だけとする。これは、話題展開権を維持できるため、細かい話題が生じにくいためである。これに対して、対話ではSE でも 2 レベル求めるが、詳細は省略する。話題スコープは、FE と同様の方法で求める。

4 話題構造認識の評価実験

人間が言語データを見て認識した話題構造と、3 章で提案した手法を用いてシステムで認識した話題構造の比較を行なった。評価用のテキスト・データとしては新聞記事 44 件を用いた。それらの記事には 488 文、単文数にして 1,726 単文が含まれている。ここで、単文とは、1 つの述語だけを持つ単位である。モノローグ・データとして用いたテレビ・ニュースの原稿 42 件には、376 文、単文数にして 2,103 単文が含まれている。対話データには 639 単文が含まれており、時間にすると約 30 分の長さである。

話題語と話題スコープに関して、適合率と再現率を求めた。人間が認識した話題構造を M、システムが認識した話題構造を S、それらの共通部分を I とすると、適合率は I/S であり、再現率は I/M である。話題スコープに関する適合率と再現率というのは、話題スコープの長さに応じて重み付けが行われている。例えば、長い話題スコープを持つ話題語に対して認識誤りを起こせば、短い話題スコープの話題語に対する誤りよりも、適合率や再現率を顕著に落とすことになる。

評価結果を表 4-1 に示す。話題スコープに関する適合率と再現率は、対話では 56.5%と 73.0%であり、モノローグでは 64.9%と 68.8%であり、テキストでは 72.1%と 75.1%である。この精度は我々の目的には十分である。というのは、映像検索システムでは話題構造を目次として用いているが、たとえ人間が認識しなかったものを計算機が話題語として誤認識しても、人間はそれが誤りであるということを推定することができるからである。

モノローグとテキストに対する適合率、再現率は、話題語に関するものよりも話題スコープに関するものの方がよかった。これは 1 つには、長い話題スコープを持った FE 話題が正しく認識されていることによる。この傾向は、目次に使うという我々の目的に適合している。

	適合率(I/S)		再現率 (I/M)	
	話題語	話題スコープ	話題語	話題スコープ
対話	(64topics/100topics) ×100 = **64.0%**	(1375ss/2434ss) × 100 = **56.5%**	(51topics/63topics) × 100 = **81.0%**	(1375ss/1884ss) × 100 = **73.0%**
モノローグ	(109topics/214topics) ×100 = **50.9%**	(2267ss/3493ss) ×100 = **64.9%**	(109topics/166topics) ×100 =**65.7 %**	(2267ss/3295ss) ×100 = **68.8%**
テキスト	(101topics/149topics) ×100 = **67.8%**	(2115ss/2934ss) ×100 = **72.1%**	(101topics/159topics) ×100 = **63.5%**	(2115ss/2815ss) ×100 = **75.1%**

"topics" は話題語を、"ss" は単文を示す。.

表 4-1　話題構造認識実験の結果

対話においても長い話題スコープのFE話題は正しく認識されているが、精度は話題スコープに関するものよりも、話題語に関するものの方がよい。この原因の1つは、システムが話題スコープ、すなわち話題のレベル関係の認識で誤りが起きていることである。例えば、「次に」という語は、必ずしも話題転換の手掛かり句として用いられるわけではなく、「次に来る電車は」のように話されている内容における順序関係を表す場合や、何度も繰り返し発話されるような場合もある。本システムには、これらにある程度対応するための規則が組み込まれているが、それらでは対応しきれない場合もある。

我々の話題構造認識手法の問題点の1つは、話題終了の検出が困難であることである。手掛かり句は話題開始、転換、入れ子終了のいずれも基本的には、新話題の開始を示すものであり、終了は分からないので、基本的には同レベルの話題が来るまでは継続しているものとしている。。また、SE話題の認識規則でも同様で、次の話題が来るまで、その話題は継続するとしている。ただし、この問題は目次として利用する場合には、重大ではない。

5 まとめ

話題構造を用いて目次と索引を提供する映像検索システムを提案した。これらの機能により、人間にとっての情報圧縮が可能とあることが確認された。さらに、このシステムで用いる話題構造認識手法を提案した、この手法は、「HCにおいて意味的情報を伝達にするためには話題構造を伝える必要があり、そのためには言語的手掛かりが用いられたり、言語現象が伴われたりする」という仮定に基づき、対話、モノローグ、テキストという3種類の伝達形態に応じた手掛かりを規則化して認識を行なう。また、評価実験によって本手法によって、十分な精度が得られることが確認された。

6 参考文献

[大久保94] 大久保雅且、中川透「AV情報構造化技術とその情報要約への応用」, 情処情報メディア研究会(IM15), 1994.

[竹下93] 竹下敦「話題構造認識の観点からのヒューマンコミュニケーションの研究」, 信学会 秋季大会 D-62, p.6-64, 1993.

[坂本91]坂本純次、平社豊「文字放送デコーダ用LSI開発、米国でのテレビ内蔵化法案に対応」, 日経エレクトロニクス, p.149-158, 1991.

[Cohen87] R. Cohen. Analyzing the structure of argumentative discourse. *Computational Linguistics*, 13:11-24, 1987.

[Danes74] F. Danes. Functional sentence perspective and the organization of the text. In F.Danes, editor, *Papers on functional sentence perspective*, pages 106-128. Academia, 1974.

[Grosz86] B. Grosz and C. Sidner. Attention, intention and the structure of discourse. *Computational Linguistics*, 12(3):175-204, 1986.

[Mann90] W. Mann and S. Thompson. Rhetorical structure theory: Description and construction of text structures. In G.Kempen, editor, *Natural Language Generation*, pages 85-96. Martinus Nijhoff, 1990.

[Oviatt89] S.L.Oviatt and P.R.Cohen. The effect of interaction on spoken discourse. In *ACL-89*. pages 126-134, 1989.

[Schegloff79] E. Shegloff. Identification and recognition in telephone conversation openings. In G. Psathas, editor, *Everyday language: Studies in Ethnomethodology*. Irvington, 1979.