

# 計算機マニュアル文を対象とした英日機械翻訳に 関する一考察

堤 泰 治 郎

(日本アイ・ビー・エム(株) 東京サイエンスワーク・センター)

## 1. はじめに

機械翻訳において対象分野を限定しさらに翻訳の言語対も固定することにより、高品質の翻訳が期待できる<sup>1)</sup>。筆者は、現在計算機マニュアル文の英日機械翻訳について研究を行っており<sup>2)</sup>、特にいかにして人間の翻訳者による訳文に近い自然な翻訳結果を得るかに重点をおいている。

本稿では、計算機マニュアル文の分野を限定した場合の英日機械翻訳について考察を行う。一般に、実用的な翻訳システムを作成するためのポイントとしては、

(1) 翻訳者の翻訳技術を可能な限り機械化する

(2) 翻訳者が使用する翻訳辞書と同程度の情報が利用可能な機械辞書を作成する

などが考えられるが<sup>3)</sup>、ここでは上記のことから実現するための具体的な手段について述べる。2章では、計算機マニュアル文の特徴と計算機処理上の英日比較について概説する。これらの結果は、分野限定により英日比較がより明確に整理されうることを示している。3章では、英日翻訳システムの概要を述べ、計算機マニュアル分野における名詞の意味マーカとこれらを直接的に利用した各種の変換辞書について説明する。4章では、変換部分を一部LISP上でのインプリメントした結果について述べ、本方式によって翻訳者によるような自然な訳文が得られる可能性を示す。

## 2. 計算機マニュアル文の特徴と英日比較

調査対象としたマニュアルは、IBM Introduction to DOS/VSE, DOS/VSE System

Utilities, VM/SP General Information Manual などである。今回は手作業により定性的な特徴を把握した。

### 2.1 計算機マニュアル文の特徴

調査の結果、次のような特徴が明らかになった。

(1) 用語の範囲が限定されている。たとえば、動詞 --- 数百種程度、形容詞・形容動詞 --- 数十種程度、副詞 --- 数十種程度、名詞 --- 基本名詞(1~2万種類、複合名詞・固有名詞(常増加)など)である。これらについては現在統計的に調査中である。

(2) 動詞についての格構造が明確で単純である。たとえば、会話文に現われるような感情を表わす動詞や文修飾副詞が現われない。

(3) 構文がある程度限定されている。たとえば、(i) 動詞の種類・用法の限定、(ii) 慣用句の限定、(iii) 係り受けの限定、(iv) 副詞の位置など。

(4) 文の長さが比較的短かく、明確な表現が多い。たとえば、複文の埋込みのレベルはせいぜい2レベルまでである。

(5) 長い名詞句が多く、また特に名詞句の並列句が多用されている。

(6) 出現する文のタイプとしては、疑問文、命令文はごくまれで平叙文、表題文(名詞句)がほとんどである。

(7) 指示詞が指示するものがかなり明確である。ほとんどの場合、その文中かまたは一つ前の文中にその指示するものが存在する。

### 2.2 計算機マニュアル文の英日比較

以下に主な比較項目について概説する。

#### (1) 関係代名詞

英日対比で考慮すべき点は、(i) 制限用法があるか非制限用法があるか。

(iii) 主文の長さとの埋込み文の長さとのバランス。(iv) 主文の時制・アスペクトとの埋込み文の時制・アスペクトとの関係などである。調査したマニュアルの場合、非制限用法のときは常にコンマが挿入されていて、この場合は2文に分割して訳出する必要があるが、その他の場合には埋込み文を先に訳す文型で1文として訳出して問題はない。

(2) 限定語・限定補足語・否定辞・モーダルを表わす語句

一般には、限定語、限定補足語、否定辞、およびモーダルを表わす語句の組合せによって一つの概念が決定される。日本文への変換の場合には、これらのものを個々に変換するのではなく、一つの組合せとして同時に変換しなければならない。これらは、日本文での連体詞・形容詞・形容詞相当語句(各、あの、あらゆる、一つの、...)、係助詞・副助詞(も、さえ、のみ、ずつ、...)、および否定表現に關係する。表1にこれらの例を示す。

	限定語 (名詞句を限定)	限定補足語 (限定語と同時に用いてそれを補足)	否定辞 (述語否定)	モーダルを表わす語 (助動詞等によるモーダルの表現)
限 語	a(an) the	also only	not(助動詞の 直後に使用され るもの)	can may
	any(of) each(of)	but alone		must should have to need(助動) may have to
	every all(of)	always		had better
	some(of) many(of)	especially particularly principally		require(動) need(動) necessary(%)
	half(of) double couple			

表1. 限定語・限定補足語・否定辞・モーダルを表わす語

この英日比較は、英語と日本語の大きな違いの一つに關係するものであり、現在詳細な英日比較について検討中である。

(3) 指示代名詞・人称代名詞

表2に計算機マニュアルでの例を示す。

代名詞	指示物の代表例
he	オペレーター
they	プログラム類
you	ユーザ
{it, this, that} these, those	(多様な使途)

表2. 計算機マニュアルでの代名詞

特徴的なのは、they = プログラム類、you = ユーザの対応がほぼ100%であることである。また、これら以外の it, this などについては訳出のルールは表3のようによまとめることができる。

条 件	訳 出 ルール
・日本文で主語となり、かつ、指示するものが明確	・省略する
・指示するものが明確	・直訳("それ、...")する
・日本文の場合で英文で前方照応が日本文で後方照応の状況になる	・指示されるもので復元し、かつ、本来の指示されるものを省略したり、指示語+基本名詞
・特に固有名称が代名詞化されている	・指示されるもので復元

表3. 代名詞訳出ルール

(4) 副詞

出現する副詞はほとんど命題内副詞<sup>(4)</sup>である。(i) 時制・アスペクトの副詞(already, later, ...), (ii) 場所の副詞(here, ...), (iii) 頻度の副詞(always, frequently, ...), (iv) 強意・程度の副詞(additionally, considerably, ...), (v) 様態の副詞(immediately, successfully, ...)などがある。訳出ルールとしては、NP<sub>1</sub>, NP<sub>2</sub>を名詞句、Adv.を副詞、Vを動詞として、  
 (a) NP<sub>2</sub>の長さが短い --- 「NP<sub>1</sub>はAdv., NP<sub>2</sub>はV<sub>1</sub>する」  
 (b) NP<sub>2</sub>の長さが長い --- 「NP<sub>1</sub>は, NP<sub>2</sub>はAdv., V<sub>1</sub>する」  
 (c) Adv.が文頭にある --- 「Adv., NP<sub>1</sub>はNP<sub>2</sub>をV<sub>1</sub>する」  
 であるが、(c)の場合以外は、(b)のように訳出してほとんど問題はない。

(5) 「動作名詞 + of + 名詞句」の型

英日変換で特徴的なのは変換の一つがこの型の名詞句の変換である。すなわち日本文でも同様に名詞句として訳すかまたは叙述表現として訳すかの選択の問題である。たとえば、「動詞 + 動作名詞 + of + 名詞句」の型の文の場合、この「動詞」の種類によって次のように選択ができる。

- (a) 「動詞」が is, offer, facilitate, make (use) などの場合 --- 叙述的に訳出。
- (b) (a) 以外の場合 --- 名詞句として訳出。

(6) 時制・アスペクト

時制・アスペクトについては、一般に英語の方が日本語よりも表現が豊かであるのび英日翻訳の場合には特に大きな問題はないと考えられる。ただし、時制・アスペクトに影響を与えるもの、

たとえ、begin, keep などの動詞や副詞、また複文における主文、従属文の時制・アスペクトのバランスも考慮されなければならぬ。表4に単文レベルでの英日対比表を示す。

英語(日本語)		時制				アスペクト	
時制	アスペクト	現在	過去	未来	現在進行形	過去進行形	未来進行形
PRES	ACT	-	-	-	ている	-	-
*	PROG	いる	-	-	いる	-	-
*	PERF	した	した	した	している	していた	している
*	PAS	-	-	する	している	-	-
*	PROG	いる	-	-	いる	-	-
*	PERF	した	した	した	している	していた	している
PAST	ACT	-	した	-	していた	した	した
*	PROG	いた	-	-	いた	-	-
*	PERF	した	した	した	していた	していた	していた
*	PAS	-	-	する	している	-	-
*	PROG	いた	-	-	いた	-	-
*	PERF	した	した	した	していた	していた	していた
FUTU	ACT	-	-	する	している	する	する
*	PROG	いる	-	-	いる	-	-
*	PERF	した	した	した	している	していた	している
*	PAS	-	-	する	している	-	-
*	PROG	いる	-	-	いる	-	-
*	PERF	した	した	した	している	していた	している

表4. 時制・アスペクト対比表

(7) 文脈処理

計算機マニユアル文において一意に文意を決定できない原因としては、

- (i) 多品詞語、同形異義語 および 多義語に基づくあいまいさ
- (ii) 係り受けのあいまいさ
- (iii) 代名詞化、省略によるあいまいさ
- (iv) 文体(視点)の決定要因のあいまいさ

などがある。これらのかみりのものは、当処理文の範囲内の統語処理、意味処理で解決可能であるが、前文にさかのぼっての文脈処理、背景の知識を利用した処理も必要になる場合もある。

ここでは、alsoに関する文脈処理の例を示す。表5に計算機マニユアルの調査結果を示す。表5に示すように、頻度40%を占めるタイプIV, V, VIについては統語処理により容易に確定できるが、残り60%のタイプI, II, IIIの場合には、alsoがどこに係るかは以前の文も取扱わなければ決定できない。この場合、簡単な処置として、II, IIIのタイプに対してもタイプIのように「も」をV1のあとに位置づけて、その「も」が

C1, C2, V1のいずれにも係りうるように故意にあいまいな表現ひきまわすのが文脈上奇妙な日本語文章になっってしまうことが多い。

英語	文中の位置	出典	タイプ	文	出典	出現頻度	割合	
adv.	副詞	I	I	C1 also V1 C2 C3 ...	C1 ... V1 ...	610	41%	
				II	C1 also V1 C2 C3 ...	C1 ... V1 ...	27(17)	(25%)
				III	C1 also V1 C2 C3 ...	C1 ... V1 ...	<376>	<40%>
adv.	副詞	II	II	C1 also V1 C2 C3 ...	C1 ... V1 ...	114	15%	
				III	C1 also V1 C2 C3 ...	C1 ... V1 ...	<156>	<15%>
adv.	副詞	M	M	C1 V1 also C2 C3 ...	C1 ... V1 ...	10	<11%>	
				(C1 also C2 ...)				
adv.	副詞	V	V	Also, Sub-conj ...	主文 ...	8	2%	
				(Also, S ...)			<114>	<10%>
adv.	副詞	M	M	C1, C2, and also C3 ...	主文 ...	11	<15%>	
				V1, also V2 ...				

表5. 計算機マニユアル文における'also'の出現結果

alsoを含む文に対してはそれ以前に必ずそれと対照すべき文がある。C1, C2, V1の共通部分、および異なる部分を調べればタイプI, II, IIIの区別は可能になる。実例について調べた結果、約60%が決定可能であることがわかった。

3. 計算機マニユアル文を対象とした英日機械翻訳システムについて

ここでは、計算機マニユアル文の調査結果をもとづいて考えた英日翻訳システムについて述べる。

3.1 概略と特徴

図1に翻訳プロセスの概略を示す。

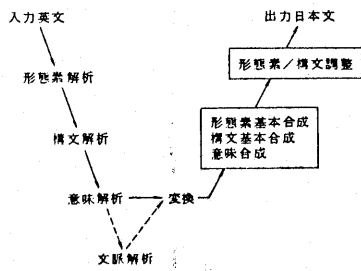


図1. 英日翻訳プロセスの概略図

このシステムの特徴は次のとおりである。

- (1) 名詞の意味マーカを直接的な手段として、解析された英文の中間表現から一挙に表層に近い日本文へ変換する。
- (2) 英文の表層型、語彙情報を保存し、それらに対応する日本文の適切な文型、語彙として変換することにより文体の処理を簡単にする。
- (3) 文の変換は、英文の1組の名詞、動詞、助動詞、前置詞などを日本文の1組の名詞、動詞、助動詞、助詞などに同時に変換するのを基本とする。また、副詞、限定詞、否定語、時制・アスペクト・モダルについても同様である。
- (4) システムの中核に辞書、変換・変形テーブルを位置づける。
- (5) 英文解析は、基本的に統語情報により行い、複数の解析木の意味的な選択は変換時に行う。

図2、図3に英語、日本語それぞれの中間表現を示す。どちらも句構造表現を基本としている。

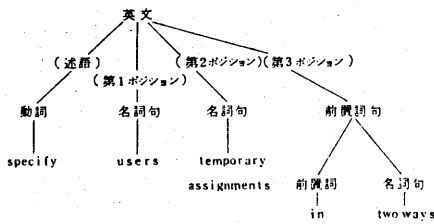


図2. 英語中間表現の例

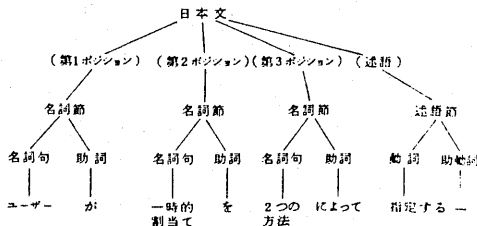


図3. 日本語中間表現の例

この日本語中間表現では、すむに基本的な構文合成、形態素合成は完了しており、最終的に表層文をうるには、がは変形、名詞節相互の位置の変換などを行えばよいような表現となっている。

### 3.2 名詞の意味マーカ

計算機マニュアル分野における名詞の意味マーカを次のように設定した。

- (1) 論理的容器 (LC) - 情報を蓄えておく論理的な場所。(area, file, library, buffer, ...)
  - (2) 論理的内容物 (LE) - 論理的容器に蓄えられる情報。(data, record, transaction, string, ...)
  - (3) マニュアル・メモ・レター (DM) - 文書そのもの。(manual, guide, material, document, ...)
  - (4) システムの状態 (ST) - 計算機システムの各部または全体がとりうる状態。(interrupt, overflow, exception, active, ...)
  - (5) 技術・手法 (TH) - 計算機システムを効果的に運営するための個々の具体的な技術・手法 (buffering, dispatching, ...)
  - (6) 機能・能力 (FT) - ある技術・手法により可能になる機能・能力 (support, facility, function, ability, feature, ...)
  - (7) 知識・情報 (IF) - 人間、機械が参照する情報 (information, introduction, ...)
  - (8) 属性 (FA) - 計算機システムの各部を説明したり評価したりする場合に使用される評価尺度 (interval, time, pitch, level, ...)
  - (9) 属性の実現値 (FV) - (8)の属性に対する値。(50 bytes, 10 seconds, "ADD", ...)
  - (10) システム関係者 (SP) - 計算機システムに関係する人間 (operator, programmer, ...)
  - (11) 装置・機器 (UD) - ある物理的な装置・機器 (station, unit, display, clock, ...)
  - (12) 作業 (WK) - 計算機システムに関連する各種の作業 (test, initialization, restart, ...)
  - (13) 用具・用品 (SL) - 計算機システムの消耗品など。(card, diskette, disk pack, ...)
  - (14) 論理的伝達路 (LP) - 情報・コントロールが伝わる通路。(data bus, network, line, ...)
- そのほか、一つの叙述のまとまりである叙述 (PS) やそれについての属性を表わす叙述属性 (PA) (objective, purpose, ...) などもある。

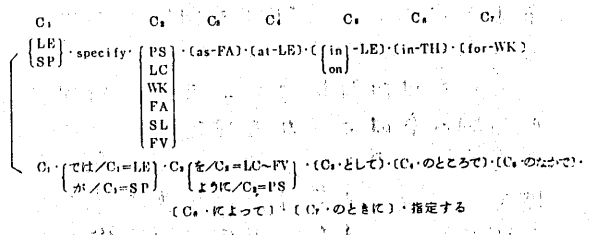
### 3.3 変換辞書

本システムでは、いわゆる格構造<sup>5)</sup>による方法は採用せず、ここで提案した名詞の意味マーカを直接の手がかり

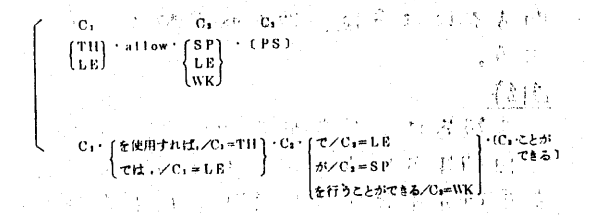
リとして翻訳のための変換辞書を効果的に構成することになっている。変換辞書のうち特に重要なものは動詞の変換辞書であり、高品質の翻訳を行う場合の大きなポイントになる。すばりに述べたように計算機マニュアル分野では動詞の種類は数百種類があると考えられるので、個々の動詞について詳細な検討を行い、辞書を構成していくことに特に問題はないと考えられる。

(1) 動詞の変換辞書

次に specify についての記述内容を示す。



ここで、C<sub>1</sub>~C<sub>7</sub>は、specifyを中心とする構文上の位置を示す。{}は選択、[]は省略可、as, at, in, ... は前置詞を表わしている。もう一つの例として allow についての記述例を挙げておく。



英日の表現の対比のうちで、すれの大きな現象として使役、能動/受動、自動/他動および時制、アスペクトに関することがらがある。こゝでの方式によれば、個々の動詞の性質に応じて、上記の現象に関するいかなる変換・調整も可能になる。小ルールの組合せで実現する方法に比べて“例外”がなく、また一部の修正による副作用もないために維持・管理が容易になる。特長をまとめると次のようになる。

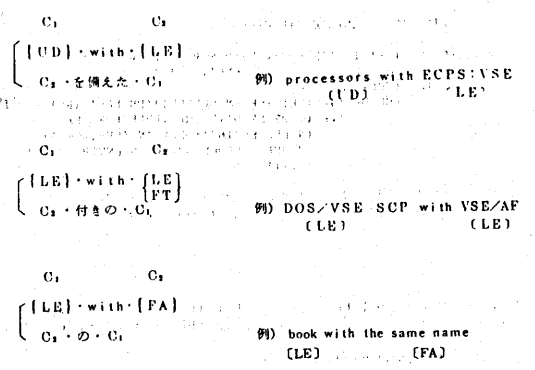
- (i) ある動詞を中心とする基本的な文

について英日変換にともなうその動詞特有の現象をそのまま記述できる。

- (ii) 基本的な文型の英日の対応も同時に表現できる。
- (iii) ある動詞とそれと同時に使われる前置詞(前置詞句)が特別の意味を表わす場合が多くあり、このような現象もそのまま記述できる。
- (iv) 人間が参照する英和辞書の主な情報を再整理した形で保有している。

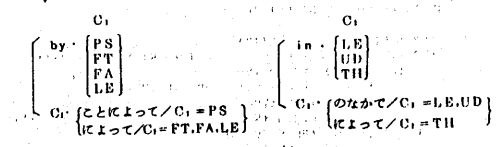
(2) 「名詞句 + 前置詞句」の型の名詞句のための変換辞書

計算機マニュアル文では、短い文で簡潔に記述するという指向からこの表現は多く用いられている。これらの名詞句は一つのまとまった意味をもち、意味マーカーによって訳し分けが可能になる。次に、with についての例を挙げる。



(3) 前置詞句の変換辞書

前置詞の訳出の場合、動詞に強く依存するものは、動詞の変換辞書に記述しておけばよい。それ以外のいわゆる自由格となるものについては別設けた前置詞句変換辞書を使用する。次に by と in の例を挙げる。



変換辞書としては以上のほかにも、be 動詞、副詞、接続詞のためのものなどがある。また時制・アスペクト・モダリティなどの変換についても2.2 (6)に述べたことを基本とした変換テーブルを使用する。

#### 4. 英日変換部分の実験

3章の内容にもとづいて、英日翻訳過程のうちの変換部分を一部 LISP上 にインプリメントし、提案した各種変換辞書の有効性を検証した。入力的前提としては、英文の解析、変形を終了し、英語中間表現を得た段階としている。また、狭義の名詞句についての訳出は独立した処理を予定しているのもそれらもすでに訳出されて意味マーカ付けられていておられるものと仮定している。次に、実験結果を示す。

- The-DOS/VSE-SCP with VSE/Advanced-Functions provides the-support needed for processing in a-multi-programming-environment.

<ENGLISH INTERMEDIATE STRUCTURE> :

```
(S (NP (DET THE) (N DOS/VSE) (N SCP) (PRPP WITH (NP (A VSE/ADVANCED) (N FUNCTION))))
  (VP (V001 PROVIDE)
    (VCOMP (NP (DET THE) (N SUPPORT) (PS-PART (NP (VP (V002 NEED) (VCOMP (PRPP FOR (NP (N PROCESSING) (PRPP IN (NP (DET A) (N MULTI-PROGRAMMING) (N ENVIRONMENT) )))))))))))
```

VSE-カクショウ・キノヲソナイ DOS/VSE-SCP ハ、ガシユウ・フ・ロク・ラミンク・カンキョウ ニ オカク シヨリ ニ ヒツヨク ナ ヲホ・ト ロソナイ タイマス。

- Messages received from remote-terminal are processed by an-application-program running in the-central-processor.

<ENGLISH INTERMEDIATE STRUCTURE> :

```
(S (NP (N MESSAGES) (PS-PART (NP (VP (V001 RECEIVE) (VCOMP (PRPP FROM (NP (A REMOTE) (N TERMINAL))))))
  (VP (V002 PROCESS) (VCOMP (NP (PRPP BY (NP (DET AN) (N APPLICATION) (N PROGRAM) (PS-PART (NP (V003 RUN) (VCOMP (PRPP IN (DET THE) (ADJ CENTRAL) (N PROCESSOR))))))))))
```

インカク・タンマツ カラ ハイチ クル ヲホ・シ ハ、 ヲウオウ・インサン・ショリソクテ ショウコク シルテイル テキョウ・キョウム・フ・ロク・ラム ニ ヨッテ シヨリ シルマス。

- The-facility allows your-programs to process data that is stored in fixed-block-DASDs, although DOS/VSE is operating in System/370 mode.

<ENGLISH INTERMEDIATE STRUCTURE> :

```
(S (NP (DET THE) (N FACILITY))
  (VP (V001 ALLOW) (VCOMP (NP (ADJ YOUR) (N PROGRAMS))
    (PRP-PS TO (PS-TO (NP (VP (V002 PROCESS) (VCOMP (NP (N DATA) (PS-REL (NP (VP (V003 STORE) (VCOMP (PRPP ON (NP (ADJ FIXED) (N BLOCK) (N DASDS)))))) ))))))))
  (SUBC ALTHOUGH (PS-SUBC (NP (N DOS/VSE))
    (VP (V004 OPERATE) (VCOMP (PRPP IN (NP (N SYSTEM/370) (N MODE))))))
```

コノキノヲ シヨクシルハ DOS/VSE カ SYSTEM/370-モード テ カトウ シルテイル ハ・アイデ・モ・ユウ・フ・ロク・ラム ニ ヨッテ コテイ・フ・ロク・ネウキノ・DASD ニ キョク シルテイル テ・クワ・ヲ シヨリ スル コト ガ テキマス。

#### 5. おわりに

本報では、計算機マニュアル文分野を限定した英日機械翻訳システムについて考察した。今回は手作業によるマニュアル文の調査のみであったが、対象とするマニュアル文を十分に調査検討すれば、言語現象のかなりの部分について境界が明確になり、計算機処理に適した形で効果的に整理できるとい見通しを得た。これにより、範囲は限定されるが、工夫次第で人間によるような高品質の翻訳も期待できることがわかった。

ここでは計算機マニュアル分野固有の名詞の意味マーカを設定し、それらを直接に利用した各種の変換辞書、変換テーブルが重要な役割を果たすことを述べた。また、実際の計算機マニュアル文について英語中間表現を仮定し、それを変換部の入力として行った実験の結果について述べた。

この方法がどの範囲まじの計算機マニュアル類に対して有効であるかを判断するには今後、多くの実験を必要とする。

#### 謝辞

本報告は、筆者が昭和56年4月より昭和57年3月まで京都大学電気工学科長尾研究室に依託研究生として在籍し研究を行、其結果をまとめたものである。上記期間中御指導いただき貴重な御示唆を下さった長尾真教授に深く感謝いたします。

#### 参考文献

- 1) ソフトウェア産業振興協会：技術文書翻訳支援系 (1979) p27-47
- 2) 堀：計算機マニュアル文を対象とした英日機械翻訳に関する一考察。東京工科大学リサーチ・レポート N:G318-1571 (1982)
- 3) 片桐恭弘・野村浩郎：英日機械翻訳における意味構造変換に関する一考察。情報処理、自然言語研究 24-2。
- 4) 国広哲爾編：日英語比較講座 第二巻・文法、大修館(1980) p191-219
- 5) Fillmore, C. J. : The Case for Case; Universals in Linguistic Theory, Holt, Rinehart & Winston, INC (1968)