

パソコン通信を利用した機械翻訳電子会議システム

西野文人 伊吹潤 中村直人
富士通研究所

我々は機械翻訳システムの様々な利用方法を検討している。現在の機械翻訳システムは限られた種類の文章を大量に翻訳するために開発されたものであり、有効に利用できる場面は限られている。我々は機械翻訳システムの利用分野を広げる試みの一つとしてパソコン通信を利用した機械翻訳会議システムを作成した。本システムはパソコン通信の電子会議室に一利用者として機械翻訳システムが参加し、人の発言を読むと自分の発言として翻訳結果をアップロードすることによって複数の言語間での会議の補佐をするものである。本稿では本システムの構成と特徴について述べる。

A MACHINE TRANSLATION CONFERENCE SYSTEM USING PERSONAL COMPUTER TELECOMMUNICATIONS

Fumihito NISHINO, Jun IBUKI and Naohito NAKAMURA

FUJITSU LABORATORIES LTD.
1015, Kamikodanaka, Nakahara-ku, Kawasaki 211, Japan

This is a report on a machine translation conference system for personal computer telecommunications. Although various machine translation (MT) systems are now available, their use is limited only to professional users due to restrictions on sentence patterns and vocabulary. To lure personal users to MT systems, we are now developing an MT conference system for PC telecommunications, where an MT system serves as a simultaneous interpreter at a conference. In this paper we describe the basic configuration of the system, and then discuss the characteristics of PC telecommunications and translation problems associated with it.

1. はじめに

これまで機械翻訳はマニュアルや論文などを対象として、翻訳を専門とする人が工数を削減できるようにするためにはどうしたら良いかという観点から開発されてきた。その結果、現在の翻訳品質は十分とは言えないが、ある程度のレベルに達し、多くの実用システムが商品化されるにいたっている。しかし、まだまだ機械翻訳は非常に限られた使い方しかされていない。現在の翻訳品質を前提として、機械翻訳システムの利用のしかたを研究していくことが課題となっている。一方、通信機能をもったワープロやパソコンの普及に伴って、パソコン通信利用者が増加しており、電子メールのやり取りや電子会議室での議論などが行なわれている。そこで、我々はパソコン通信と機械翻訳システムとを結合することで、機械翻訳の新しい利用方法を検討している。その一つとして、電子メールを使用することにより、誰でも手軽に機械翻訳が使用できるシステム[1]の開発を行なった。そして今回さらに、機械翻訳の別の利用法として、パソコン通信の電子会議室を使用した機械翻訳システムを試作した。これは、機械翻訳システムも一利用者としてパソコン通信に参加し、人の発言を読むと、自分の発言として人の発言の機械翻訳結果をアップロードすることによって実現したものである。本稿では、本機械翻訳電子会議システムの概要と問題点について述べる。

2. パソコン通信とは

我々はNIFTY-Serveというパソコン通信システムを利用して本システムを実現した。本節ではNIFTY-Serveにおける電子会議室について述べる。

商用のパソコン通信であるNIFTY-Serveではテーマごとにフォーラムとして別れている。そして、そのフォーラムの中に、電子会議室、電子掲示板、リアルタイム会議室などが存在する。

電子会議室はさらにその中がいくつかのセクションに分離しており、各セクションでは、利用者が自由に発言したり、前に誰かが発言した内容を読んだりすることができるようになっており、様々な議論やおしゃべりがおこなわれている(図1)。

-MES(8)翻訳実験会議室(日本語発言用) 90/04/09
019/019 PDG00555 神田 泰典 すこし修正をお願いします
(8) 90/04/08 12:11

これまでの、8と9の会議室での実験で大体のことは理解できました。そこをお願いですが、半角と全角の問題や空白で文章が切断されてしまって、エラーになるのを修正してもらえませんか？これがなれば、それなりに使えるような気がします。また、ネットワークでの発言の体裁にたいしての反応もわかりますので、これを少し研究して対応できるようにすれば実験をもっと推進できると思います。長い文章でもまともに翻訳されている例もあり、それなりに使えると思います。

図1 電子会議室発言例

一般の電子会議室では発言の時刻と読者がその発言を読む時刻との間にはタイムラグがあるが、リアルタイム会議室は、電子会議をリアルタイムに行うものである。すなわち、会議室内で誰かが発言すると、その内容は発言者のハンドル名とともにただちにその時点で会議室に入っている他の利用者全員に表示される(図2)。

(西野)クリフさん、読めますか？
(クリフ)はい、私は読むことができます。
(西野)昨日はどうでしたか？

図2 リアルタイム会議室発言例

3. パソコン通信と機械翻訳との結合

パソコン通信の中にはオンラインショッピングやデータベースサービスなども提供されていることもある。しかし、これらのシステムはあらかじめパソコン通信システムのホストに組み込まれているものである。したがって、一般の利用者がそのような機能を自由に追加したり、変更したりすることはできない。一般利用者は与えられたメニューに従って電子会議室やリアルタイム会議室で発言をしたり、人の発言を読んだりすることができるだけである。

しかし、このような制限下においても、一利用者として自動通信プログラムを用いて、自動

的に会議室の発言をダウンロードしたり、アップロードすることができる。そこで我々は、UNIX上から使用できる自動通信プログラム言語を作成し、これを使用してパソコン通信のコマンドの制御やUNIX上のプロセスの制御を行なうことにした。これによって、我々はこれらのパソコン通信システムの一利用者として機械翻訳を位置づけ、他の人の発言を読むとその発言を機械翻訳した結果を自分の発言としてアップロードすることによって機械翻訳会議室を実現するシステムを作成した(図3)。以下ではまずバッチ型の機械翻訳会議室について述べ、次にリアルタイム型の機械翻訳システムおよびリアルタイム型の機械翻訳会議室について述べる。

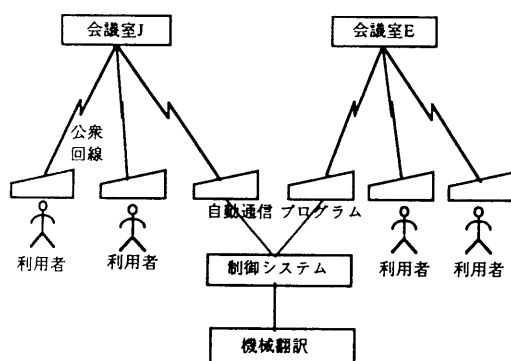


図3 パソコン通信を利用した機械翻訳会議システム

3.1. 機械翻訳会議室

複数の電子会議室間を機械翻訳システムで結合することにより機械翻訳会議室が実現される。すなわち、会議室のあるセッションJの、利用者は日本語で発言し、日本語の発言内容を読む。これに対して、もう一つセッションEを設ける。このセッションEでは、利用者はセッションJで発言された内容の英語に翻訳された結果を読む。そして、セッションEに英語で発言したものは機械翻訳会議システムによって日本語に翻訳され、セッションJにも登録される。なお、このセッションJとEは同一フォーラム内にあってもいいし、異なるフォーラムあるいは異なるパソコ

ン通信システムであってもよい。

上記機能を実現する機械翻訳会議システムは、パソコン通信制御部と機械翻訳部とからなる。パソコン通信制御部は、何分かおきにパソコン通信システムの中に入り、指定されたセッションに行く。そして、システムが未読な記事がないかどうかチェックする。もし、未読な記事があれば、記事ごとにダウンロードし、発言者、標題、本文を切りだす。これらはそれぞれ機械翻訳システムによって、それが日本語のフォーラムのものであれば英語に、英語のフォーラムのものであれば日本語に翻訳される。翻訳した結果は、対応する言語のフォーラムにアップロードする(図4)。これによって、日本語の会議室と英語の会議室が結ばれ、お互いに母国語で議論しあうことが可能になる。

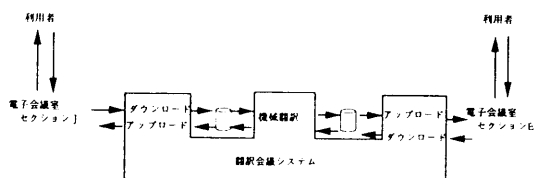


図4 機械翻訳会議システムの構成

3.2. リアルタイム翻訳

一般の電子会議室では、会議は実時間では実行されないが、リアルタイム会議室では会議は実時間で進行される。したがって、翻訳システムも発言者の発言を即時に取り込んでそれを機械翻訳してその翻訳結果をアップロードしなければならない。翻訳結果を発言者にだけ表示すると、パソコン通信を利用したリアルタイム機械翻訳システムになる。

リアルタイム機械翻訳システムは、機械翻訳を実行するプロセスとパソコン通信のリアルタイム会議室に入って利用者の発言の取り込みと翻訳結果の発言を行なうパソコン通信制御プロセスの2つからなる。前述の機械翻訳会議室では、システム側が適当な時に何か発言があるかどうか会議室に見に行けばよかったが、リアル

タイム翻訳ではシステムはあらかじめリアルタイム会議室にはいて誰かが発言するのを待ち構えていなければならない。利用者がリアルタイム会議室で発言すると、その内容は発言者のハンドル名がつけられて、その時リアルタイム会議室にいる全利用者に表示される。本システムではその表示される文字列を取り込む。そして、まずハンドル名と本文とを分離し、発言者が誰であるかを判定する。自分自身の発言（翻訳結果）でなければ、ハンドル名と本文とを機械翻訳サーバに送る。機械翻訳サーバによって翻訳された結果は、パソコン通信制御プロセスに送られる。パソコン通信制御プロセスは翻訳した結果を発言する。これによって翻訳システムのハンドル名がつけられて、リアルタイム会議室において一般利用者が自分の発言内容の機械翻訳結果を見ることができる(図5)。

これはテストです。
(西野)これはテストです。
(ATLAS)(Nishino)This is a test.

図5 リアルタイム翻訳例

3.3. リアルタイム翻訳会議

上の説明は1つのリアルタイム会議室上での機械翻訳であったが、これを発言されたフォーラムと翻訳結果を表示するフォーラムを別のものとし、このような処理を複数のリアルタイム会議室間で相互に行なうことによってリアルタイムの翻訳会議室が実施される。例えば、日本語のリアルタイム会議室と英語のリアルタイム会議室を設け、それぞれのリアルタイム会議室に1つずつプロセスがはりついて誰かの発言を待つ。日本語リアルタイム会議室で発言された文は日本語のリアルタイム会議室にはりついているプロセスJによって取り込まれ、機械翻訳プロセスに渡される。機械翻訳プロセスでは、日本語から英語へ翻訳をし、英語のリアルタイム会議室にはりついているプロセスEに結果を渡す。プロセスEは翻訳結果ファイルから順次翻訳結果を読み取って、その内容を英語のリアルタイム会議室に発言する。逆に英語会議室で発言されたものは日本語に翻訳される(図6)。

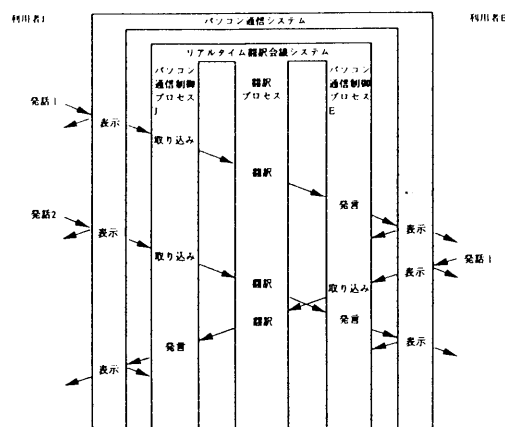


図6 リアルタイム翻訳会議の流れ

4. 文の種類

これまでの機械翻訳の多くはマニュアル、論文、特許文といった誰かがあらかじめ入力した文書を翻訳しようとするものであった。しかし、パソコン通信の会議室のようなところでの文体は少し異なる。これまで通訳を介した依頼や問い合わせに関する電話会話の分析[2]や端末間対話の言語理解方式などが検討されている[3]。しかし、実際のパソコン通信の電子会議室での会話の文の分析はあまりないようである。以下ではパソコン通信の会議室に見られる文の特徴について述べる。

4.1. 内容

マニュアルでは主に宣言文と読者に対する命令文とで、論文ではほとんどが宣言文で文章が構成されている。これに対してフォーラムの会議室中では、特定の話題に関する討論、あいさつ、記事・文献の紹介など多岐にわたるため、その文種も次に示すように多岐に分類される。実際のいくつかのフォーラムの会議室の中での特徴的な点は次のようなものである。

i) あいさつ文

会議室では、入会のあいさつ、それに対するあいさつ、質問をするときのあいさつ、回答をもらってのあいさつなど、非常に多くのあいさ

つかかわされる。例えば、以下のようなものがある。

例)

冒頭のあいさつ文

はじめまして。

XXさん、はじめまして。

XXです。

XXといます。

XXさん、いらっしやい。

お礼のあいさつ

YYありがとうございました。

結びのことば

これからもよろしくお願いします。

今後ともよろしくお願いします。

ii) 情報要求, 意見表明

会議室では、情報を要求したり、それに対する回答を示したり、自分の意見を表明したりすることが多くなる。したがって、次のような文があらわれる。

例)

情報の要求

～を教えてください。

～を教えてくださいませんか。

どうして～なのでしょう。

～をお願いします。

確認:

～のことですか?

～じゃないですか?

～ですよ。

提案:

～よいのではないですか。

同意:

確かに～ですね。

なるほど～ですね。

推測:

～と思います。

～かもしれません。

～ようです。

意見表明:

～たいのです。

会話文の種類についてはすでに[2]などで分類が行われているが、そこで対象となったのはタ

スク遂行型の会話であり、会話文の種類も限定されたものとなっている。これに対してパソコン通信の会議室内で発言は議論における主張、同意、反駁などの行為が主体となったものであり、相手と自分の意見の関係を正しく伝えることが重要となる。

4.2. 文体上の特徴

発言文は基本的に読者に向けたメッセージであり、それに起因する様々な文体上の特徴を持つ。まず文体としては口語体をとることが多い。また、発信の際の社会的な取り決め（エチケット）として敬語も多く使用される。そして、発言者の感情の表現手段として特定の終助詞、助動詞、間投詞などが多用される傾向にある。

1) 敬語表現

会議室で議論の相手に対して発言する際には、会話を円滑に行なうために、尊敬、謙譲表現がよく用いられる。

尊敬表現

～された、御～。

謙譲表現

～させていただく、伺う

2) 終助詞の多用

発言の内容に対する同意を求める、疑問の表明などに特定の終助詞がよく用いられる。

～ですか?、 ですよ。

3) 間投詞の多用

パソコン通信の会議室での発言では、単に情報を伝えるだけでなく、発言者の感情を表現する句も少なくない。実際のパソコン通信の会議室で使用されていたものの一部を以下に示す。

躊躇:

まあ、 ま、 まっ、 な～～に、

そうそう、 さあて、 うーん、 う～～ん、

うーむ、 ううむ、 えー、 え～、 えっと、

well

驚き:

あ、 あー、 ああ、 あおっ、 んん、

うーん、 む～、 おお、 やー、 うーむ、

おっと、ぎょえ、げっ、げげっ、
oh, what
読者の注意の喚起：
ほら、わい、hey
あいさつ：
ふえ～、おお、やあ、いや～、あ、
おーい、hi, hello
その他：
I say: あのおお
woops: おっと
ha!ha!ha!: あはは、にやはは、きやはは、
へへへっ

4) 省略

会話文の特徴として、「私」や「あなた」の
ような人称代名詞はほとんど省略される。

5) 記号

パソコン通信で用いられる独特な記号がある。
引用記号(>), :-), (^_^)

4.3. 翻訳処理における問題点

発言文では、字義通りに解釈した翻訳をする
と、文意が正しく伝わらない場合がある。例えば、

理由をご説明いただけないでしょうか
という文を
I wonder whether it is possible to explain reasons.
と訳すのではなく、相手への丁寧な要求である
ことを認識して、

Could you please explain the reason?
のように訳すことが必要である。

これらの文に対してはどんな種類の文である
かを理解し、さらに各種類毎に必要な情報を補
ってやらなくてはならない。その中には格関係
の補完などの会話環境の特定、発言者の態度・
感情を伝えるということも重要であり、翻訳結
果にも元の文が持っている様々なニュアンス、
丁寧さを伝えることが必要である。このため
には、上にあげた様な文体上の特徴、さらには
会話文脈を考慮した処理が必要となってくる。

なお、機械翻訳の存在を意識した利用では、
利用者は極端な口語を使うことはなく、どの
ように表現したら機械翻訳が間違わずに翻訳して

くれるかということを考えながら発言してくれ
ることをある程度期待できる。したがって、サ
ポートされていないと気になるような、簡単な
主語の補い（肯定文ならI, 疑問文ならyouを補
う）程度がしっかりされていることが重要であ
る。

5. 翻訳会議室の利用形態

翻訳会議室を考えた時に、利用者が機械翻訳
の存在をどのくらい意識するかで、次の2つの
利用形態がある。

1) 一言語中心型

現在普通に行なわれている会議であり、発言
者は機械翻訳の存在を考慮しない。したがって、
前節で述べられたような文体の特徴がある。

このような会議室での発言を機械翻訳して他
国語の人も会議に参加可能にすることは、
今後の重要な課題である。しかし、現在の機械
翻訳の実力からすると、せいぜい、他国語の人
がある程度理解できればよいというのぞき見
的な参加にとどまるであろう。

国際会議や打ち合わせのような場合には、次
のような形態を考えるべきであろう。

2) 相互乗り入れ型

はじめから機械翻訳を意識して行う。したが
って、中心となる言語が存在しない。発言者は
あらかじめ機械翻訳で翻訳されることを前提と
して発言することになる。

このような形態であれば、現在の機械翻訳で
も若干の改良と運用の仕方では何とか使うこと
ができるであろう。

6. 運用のための課題

翻訳会議室を運用して行く上では、いくつか
の課題があった。以下では個々の課題とその解
決策について述べる。

6.1 翻訳誤り文の特定・通知

会議室では、ものごとの取り決め、質問応答
など、発言内容の情報を相手に伝えることが主

目的になることが多い。あらかじめ機械翻訳の存在を意識した利用においては、発言者は比較的翻訳しやすい文章を入力してくれるが、それでも100%正しく翻訳できるわけではない。なるべく情報の伝達がスムーズに行くように支援するためには以下のようなサポートが必要である。

1) 翻訳が失敗しても何らかの情報を伝える。

これまでの翻訳家向けの機械翻訳システムでは解析が失敗したときに単にエラーメッセージだけを出力するだけで許された。というか、かえって中途半端な翻訳結果を出力するよりもよかった。しかし、翻訳会議のような応用では、解析に失敗しても、何らかの情報を相手に伝えることが重要になってくる。そのための方法としては、a)分割と、b)圧縮がある。

a) 分割

文を分割して、別個に翻訳を行なう。例えば複文を単文に分割して翻訳をする。単文の翻訳に失敗した場合には、さらにその文を句あるいは単語レベルに分割して翻訳する。

例)

原文：先回の記事で説明しておけば良かったのですが、論文の締め切りで忙しくて、すっかり忘れていました。

翻訳結果：It forgot completely the deadline of the thesis in spite of having only had to explain articles on times it was busy.

分割翻訳結果：You explain articles on times. However, I was busy in the deadline of the thesis. I forgot completely.

自立語翻訳：<<before times article explanation good thesis deadline busy forgetting>>

これらによって以下のような効果がある。

- 1) 翻訳失敗原因の局所化による誤りの局所化
- 2) 構文構造の明確化

b) 圧縮

文全体の要旨を理解する上では圧縮（修飾語句の削除）がある。例えば次のようなものである。

原文：The third, and probably the most spectacular, is marketing research, testing public receptivity to products and designs.

圧縮文：The third is marketing research.

分割や圧縮によって、伝達される情報は減るが、分割と圧縮をうまく組み合わせることで、文脈の中では、人間が読めば普通は大意は理解可能である。また、万一理解できなくても発話者に問い合わせをすればよいのである。

2) 発言者への問い合わせのための支援

翻訳失敗して簡易翻訳がなされたり、あるいは機械翻訳システムの解釈誤りによって意味不明な文に翻訳されたりした場合に、システムあるいは発言者に対して、再翻訳を依頼できるようにしておくことが重要である。システムへの再依頼としては、別解釈の表示あるいは原文の表示などがある。そのためには、どの文に対して依頼したいのかを明確に指示できるようにしておくことが必要である。また、発言者に対しても、翻訳が失敗した旨の通知、なぜうまく翻訳できなかったかをわからせるための工夫が必要である。

6.2 速度向上

バッチ型の会議室では速度はそれほど重要でないが、リアルタイム会議では、会議がスムーズに進むためには、翻訳の速度も重要になってくる。一つの発言に対して、翻訳されて相手の会議室に表示されるまであまり待たされるようであると、会議がスムーズに進行せず、利用者のイライラがつのってしまう。機械翻訳そのものの速度の向上ももちろんだが、その他の部分でいかに速いレスポンスを返せるようにするかという工夫が必要である。

1) 定型パターンの翻訳

繰り返し翻訳されるものは毎回翻訳しないようにすることで、速度の向上が図れる。例えば、リアルタイム会議室では発言者を示すために毎回発言者のハンドル名が付与されるが、一度翻訳されたハンドル名はテーブルに登録することで、繰り返し翻訳することが避けられる。

2) 文分割

文分割をしても、全体の速度としてはそれほど向上しないかもしれないが、分割された部分ごとに翻訳結果が表示されるので、議論が中断してしまうという感じが少なくなる。

6.3 前処理

情報の伝達を主目的とした文書では、情報の整理の点から個条書きやテーブル形式によって示されることが多い。また、電子会議室における特定の書式（標題、発話者など）の情報、慣用的に行なわれている書式（引用のための>など）の情報を処理したり、会話的な文に対する処理（感動詞など）することが必要である。したがって、次の2点が重要になる。

1) 文書のレイアウト構造を保存しての翻訳

2) 特殊構造の処理

これらは、機械翻訳システムの内部で処理するのではなく、前処理として独立させて実現することにより、柔軟に対処することが重要である。

6.4. スペルミスへの対処

一般の文書では、スペルミス（かな漢字変換の誤りや編集ミス、通信エラーを含む）は皆無ではないものの、比較的少ないので、これらの文書を対象とした機械翻訳では、入力文書は正しいものと仮定していることが多い。しかし、パソコン通信の会議室での発言は、一般文書と比べるとスペルミスが多い。特にリアルタイム会議室では、利用者はオンラインで入力しているために、スペルミスが非常に多く見られる。

このような会議室での翻訳は、まず情報が相手にしっかり伝わるかどうかの主眼におかれる。したがって、スペルミスをチェックして、それを修正し、機械翻訳を行なうことで、スペルミスによる機械翻訳の誤りを極小化する必要がある。

7. おわりに

パソコン通信を利用することによって我々は様々な情報を手に入れることができるようになってきた。そしてパソコン通信の世界でも国際化が進みつつあり、日本から米国のパソコン通

信にアクセスしたり、米国から日本のパソコン通信にアクセスしたりできるようになっている。また、日本国内のパソコン通信に対して英語でアクセスできる（メニューが英語で表示される）ようになってきた。これらによって、日本から米国の情報を入手したり、米国から日本の情報を入手したり、あるいは在日外国人が日本での生活情報などを容易に入手できるようになってきた。このように枠組みとしては色々整備されてきた。しかし、言語の壁が存在し、メニューは英語になっても実際の情報が日本で英語で提供される情報はまだまだ少ない。日本人が海外のパソコン通信にアクセスする場合も同様である。電子掲示板や電子会議室における日本語の情報を英語に翻訳することによってこれらが有効に活用されるであろう。このような情報は斜め読みができればいいのであり、また各個人が翻訳する必要がない。機械翻訳が有用であろう。また、これまで国際会議と言えば参加者が同一時期に同一場所に集まる必要があったわけだが、パソコン通信の発達と国際化に伴って、パソコン通信を利用した国際会議を開くということも現実的になってきた。しかし、その際、言葉の壁がコミュニケーションを阻害する要因となっていた。言語を統一することは、どうしても自由に自説を主張することが困難に感じる人が多くなるものである。ここに機械翻訳を活用することによって自由に自説を主張しあえることができる可能性がある。本稿で述べたようなシステムは今後重要な役割を果たすであろう。

謝辞

UNIXとNIFTY-Serveとの接続及びフォーラム運用にあたりご協力いただいた下浦部長(エヌ・アイ・エフ株式会社)、神田技師長、野々垣主席担当部長に謝意を表します。

参考文献

- [1]西野, 中村:「機械翻訳電子メールシステム」, 情報処理学会 NL75-5(1990)
- [2]飯田, 野垣内, 相沢:「通訳を介した電話会話の特徴分析」, 電子通信学会 NLC86-11 (1986)
- [3]小暮, 有田, 野垣内, 飯田:「端末間会話における言語理解方式」, 情報処理学会 NL62 -11(1987)