自然言語処理 72－9
（1989. 5. 19）

多言語翻訳のための中間言語の構成法

内田 裕士　　朱 美英
（財）国際情報化協力センター

日本およびインドネシア、タイ、中国、マレーシアが共同で開発している多言語間機械翻訳プロジェクトにおいて，中国と共同で研究開発中の中間言語の構成法について述べる．

多言語翻訳のための中間言語を構成するに当たっては，文章として表されている種々の情報をどのように整理し，抽象化して，各言語にあまり依存しない形で表現するのかということが重要である．本論文で提案している構成法は，これらの情報を，事象・事実，視点，意図，文章構造の４つの観点から分類刷る方法である．これにより，意味的な情報だけでなく，従来言語独立には表現しにくかった，時制，アスペクト，ムードや文体に関する情報も言語独立に表現できるようになった．

An Interlingua for Multilingual Machine Translation

Hiroshi UCHIDA　　Meiying ZHU
CENTER OF THE INTERNATIONAL COOPERATION FOR COMPUTERIZATION
5-30-9 Shiba, Minato-ku, Tokyo, 108, JAPAN

This research is about the chinese dictionary for multi-lingual machine system being developed in a cooperative research project by Japan, Indonesia, Thailand, the People's Republic of China and Malaysia.

To design an interlingua for multilingual translation, classifying and abstracting sentence information as language independent information is very important. This paper proposes a method of classifying sentence information from four points of view: evevt/fact, view-point, intention, and sentence structure. Although tense, aspect, mood and sentence style information was previously thought difficult to represent in language independent form, by this kind of classification, it is possible to represent not only semantic information but also this information in language independent form.

1. はじめに

トランスファ方式および中間言語方式の機械翻訳システムにおいては，文の解析結果を示すための中間表現が必要になる．これは，最初に文字列として与えられた文の構文構造や意味構造を示すものである．構文構造の中間表現は、木構造として表現される場合が多く，意味構造の中間表現はネットワークの形式で表される場合が多い．

トランスファ方式の機械翻訳システムで用いられる構文構造の中間表現には，入力文を解析した結果がそのまま用いられることが多いので，句構造文法を用いて解析された場合の中間表現は文の句構造を表す木構造になり，格文法を用いて解析された場合は，文の格構造を表す木構造になる．このような木構造の中間表現の場合，中間表現は原言語の構文，意味構造をそのまま反映しているので，目標言語への翻訳を行う場合には，原言語と目標言語の中間表現の間で必ず変換が必要になる．これは中間表現が言語の特徴を強く表していればいるほど，変換は大きなものとなり，多言語翻訳を目指した場合，言語対毎に必要となる変換のための作成コストが大きくなる．したがって，このような構文構造の中間表現は多言語翻訳のための中間表現としてはふさわしくないといえる．

中間言語方式の機械翻訳システムで用いられる意味構造の中間表現は，一般に意味ネットワークの形で表される．これは文の意味表現であり，文を構成する単語の表す概念が互いにどのように関係し，どのような役割を果たしているかということが表現される．このような意味構造は中間言語と呼ばれる．

機械翻訳システムを開発する上で中間言語方式は２つの重要なメリットをもっている．第１のメリットは，中間言語方式をとると，機械翻訳システムの開発を局所的にすることができるということである．ある言語の機械翻訳システムを開発するときに，言語を解析したり生成したりするための規則や辞書は，その言語を母国語とする訓練された人が開発する必要がある．中間言語方式の場合，中間言語によるインタフェースによって解析と生成は完全に分離することができ，ある言語の解析システムと生成システムの開発を他の言語の解析・生成システムと独立に進めることができる．したがって，解析あるいは生成システムの開発者はその対象となる言語と中間言語のみを知っているだけでよいことになり，機械翻訳システムの開発を局所化することができる．

第２のメリットは，機械翻訳に必要な知識を共通に使用することができるということである．人間もコンピュータ自然言語で書かれた文章を理解するためには，言葉の意味やその用法を知っていなければならない．このような自然言語を処理するための知識は非常に大規模であるが，これなくしては自然言語を人間のように処理していくことはできない．高品質の機械翻訳を行うためには，このような知識を使った意味解析は不可欠であるが，そのためには世界知識といったものが必要になる．このような知識が中間言語を用いて記述されていれば．各言語の解析システムでそのような知識を共通に使用することができ，同じ知識を別の形で開発するという無駄を省くことができるという大きなメリットがある．ただ，このようなメリット享受するためには中間言語を最適なものにする必要がある．

2. 多言語翻訳のための中間言語

中間言語は文章として表されている情報のすべてを表す必要がある．中間言語を設計するに当たっては，文章として表されている情報をどのような観点から整理し，抽象化し，意味的な表現にもっていくかということが最も重要になる．中間言語は文の意味構造を表現することを中心としたものであるが，文の意味には種々のものがある．文の表す意味のうち，いわゆる文の意味内容に関するものは比較的普遍性をもっているので，深層格を基本とした関係に基づく概念間の２項関係を表現した中間言語が広く用いられている．しかしながら，文の直接的な意味内容でない，時制やアスペクト，ムードあるいは文体によって表されるニュワンス的な意味などは言語ごとに独立に考えられ，中間言語に組み入れられていた．このような中間言語で多言語翻訳を行った場合，言語独立に考えられた情報の間での言語間変換が必要になり，中間言語方式を採ることのメリットを半減していた．また，中間言語は文の意味構造だけでなく文章の構造も表す必要があるが，これについても十分に考慮されたものはなかった．

本研究においては，従来言語に独立に考えられていた部分に重点をおき，これらの情報を抽象化し各言語に対して普遍性をもった情報として表現することを目標とした中間言語の構成法を考えた．

中間言語の基本的な構成として，中間言語の表現を事象や事実を表す情報と話者の視点を表す情報，話者や主体の意図，気持ちや判断を表す情報，文章の構造を表す情報のの４つに分けた．

事象や事実を表す枠組みとしては，EDRで開発中の概念辞書の記述枠組みと全く同一の枠組みを採ることとした．話者の視点を表す情報は，事象や事実をどこからどのようにみているのかを表すもので，過去，現在，未来など時制として表されていた情報やアスペクトとして表されていた情報を表現することとし，話者や主体の意図等に

関する情報は，事象や事実をどのような意図や判断でみる，あるいは起こそうとしているのかを表すもので，命令，疑問，断定，推量，願望などムードとして表されていた情報を中心として表現することとした．文章の構造を表す情報は，文章の構造を中間言語に反映するためのもので，文体とヘディングの関係や，前後の文との関係を中心として表現することとした．

3. 事象・事実の表現

事象や事実の表現は，EDRで開発中の概念辞書の表現と全く同一のものにした．これは機械翻訳に必要な知識を共通に使用することができるということを狙ったためである．このような知識の記述形式と中間言語を同一の記述形式にしておけば，各言語の解析システムでそのような知識を共通に使用することができ，同じ知識を別の形で開発するという無駄を省くことができるという大きなメリットがあるからである．

事象や事実の表現の中間言語の語彙は概念見出しと関係子および属性子からなる．概念見出しは，ある単語が表し得る概念を簡単な説明文で表したものである．コンピュータは概念見出しを記号として用い，それによって概念を識別し，人間は説明文で表された意味を理解し，ある概念を他の概念から識別するためのに用いるものである．この概念見出しと関係子，属性子にはEDR で開発中の概念辞書の概念見出しと関係子を採用し，それに各国語がもつユニークな概念表す概念見出しを追加して用いることとした．概念間の関係はかなり普遍的なものにすることはできるが，概念についてはこのことは当てはまらないからである．中間言語の語彙にはこうした各言語にユニークな概念も各言語に共通な概念と同様に含まれることになる．

概念間の関係を表す関係子としては，動作主(agent)，対象(object)，道具(implement) といった格関係，条件(condition) や連続事象(sequence)などの事象間関係，全体－部分(part-of)，などの意味関係，量(quant)，数(number)といった制限関係，その他，所有者(possessor)，目的(purpose) などの仮関係がある．

概念の属性を表す属性子としては，概念の集合に関する情報を与える量限定子などがある．以下に概念間の関係および概念の属性を示す．

(1) 格関係子

agent　有意志動作を引き起こす主体．有意志者，自動物が主体になる．

日本語：太郎が食べる．

c#eat －agent →c#tarou

object　動作・変化の影響を受ける対象

日本語：りんごを食べる．

c#eat －object→c#apple

属性をもつ対象

日本語：トマトが赤い．

c#red －object→c#tomato

manner　動作・変化のやり方

日本語：ゆっくり話す．

c#speak －manner→c#slowly

implement　有意志動作における道具・手段

日本語：ナイフで切る．

c#cut －implement →c#knife

material　材料または構成要素

日本語：牛乳からバターを作る．

c#butter←material－c#make

－source→c#milk

time　事象の起こる時間

日本語：８時に起きる．

c#wake-up －time→c#o'clock

－modify→c#8

time-from　事象の始まる時間

日本語：９時から働く．

c#work－time-from →c#o'clock

－modify→c#9

time-to　事象の終わる時間

日本語：11時まで働く．

c#work－time-from →c#o'clock

－modify→c#9

duration　事象の継続する期間

日本語：３時間見る．

c#see －duration→c#hour

－number→c#3

location　動作の対象となる場所

日本語：空を飛ぶ．

c#fly －location→c#sky

place　事象の成立する場所

日本語：部屋で遊ぶ．

c#play－place →c#room

source　事象の主体または対象の最初の位置

日本語：京都から来る．

c#come－source→c#kyoto

goal　事象の主体または対象の最後の位置

日本語：東京に行く．

c#go－goal→c#tokyo

(2) 事象間関係子

condition　事象・事実の条件関係

日本語：雨が降ったので家に帰った．

c#home←goal－c#return

－condition →c#rain

cooccurrence　事象・事実の同時関係

日本語：泣きながら家に帰った．

c#home←goal－c#return

－cooccurrence→c#cry

sequence　事象・事実の時間的前後関係

日本語：図書館へ行って本を借りた．

c#book←object－c#borrow

－sequence→c#go

－goal→c#library

conjunction　事象・事実間の両立関係

日本語：山は美しく，水は澄んでいる．

c#water ←object－c#clear

－conjunction →c#beautiful

－object→c#mountain

disjunction　事象・事実間の二者択一関係

日本語：学校に行くか図書館に行く．

c#library ←goal－c#go

－disjunction →c#go

－goal→c#school

(3) 意味関係子

part of　全体－部分関係

日本語：鳥の羽．

c#feather －part-of →c#bird

element of　集合－要素関係

日本語：人間の細胞．

c#cell－element-of→c#human

(4) 制限関係子

frequency　事象の起こる頻度

日本語：３回言う．

c#say －frequency →c#3

basis　基準

日本語：バラはチューリップより美しい．

c#rose←object－c#beautiful

－basis →c#beatiful

－object→c#tulip

unit　単位

日本語：１ダース当り５００円．

(c#1←number－c#dozen ←focus

－) －unit→c#yen －number→

c#500

quantity　量

日本語：３Kgのりんご．

c#apple －quantity→c#kg

－number→c#3

number　数

日本語：３Kg．

c#kg－number→c#3

modify　修飾関係

(5) 仮関係子

possessor　所有関係

日本語：太郎の犬．

c#possessor －possessor →

c#taro

purpose　目的

日本語：ゴルフに行く．

c#go－purpose →c#play

－object→c#golf

standard　比較の基準　例）彼女は私より美しい．

日本語：バラはチューリップより美しい．

c#beautiful －object→c#rose

－standard→c#tulip

degree　動作・変化の程度

日本語：３Kg痩せる．

c#become-thin －degree→c#kg

－number→c#3

and　概念間の連結関係

日本語：ローマとナポリに行く．

c#go－goal→ (－focus →c#naples

－and →c#rome)

or　概念間の選択関係

日本語：ローマかナポリに行く．

c#go－goal→ (－focus →c#naples

－or→c#rome)

(6) 属性子

all　全ての

日本語：全てのりんご．

c#apple －all →

some　ある

日本語：あるりんご．

c#apple －some→

each　各々

日本語：各々のりんご．
c‡apple －each→

4.視点の表現

話者の視点を表す情報は，事象や事実をどこからどのようにみているのかを表すものであり，過去，現在，未来など時制として表されていた情報やアスペクトとして表されていた情報が中心になっている．主に時制で表される情報は，話者が事象や事実をどの時点からみているかということを表すと考え，また，主にアスペクトとして表される情報は，話者がその時点から事象や事実をどのようにみているのかを表していると考え，その観点から整理した．

話者の視点のある時点は属性子で表され，次のようなものがある．

(1) 話者の視点のある時点

past	視点が過去	c‡pred－past→
	英語：	過去形，過去完了形
	仏語：	半過去形，大過去形，単純過去形，前過去形
present	視点が現在	c‡pred－present →
	英語：	現在形，現在完了形
	仏語：	現在形，複合過去形
future	視点が未来	c‡pred－future→
	英語：	未来形，未来完了形
	仏語：	現在形，単純未来形，前未来形

話者がある時点から事象や事実をどのようにみているのかを表す情報は，いわゆる相（アスペクト）情報についても表現をするものである．このような情報を表すために，新たに以下のような概念を導入した．

(2) 相情報を表すための概念

c‡begin	開始するがという概念
c‡end	終了するという概念
c‡continue	開始してからまだ終わっていないということを示す概念
c‡state	達成された状態や結果が残っているということを表す概念
c‡yet	まだ始まっていないということを表す概念
c‡already	すでに起こったということを表す概念
c‡soon	間もなく起こるということを示す概念
c‡just	ほんの少し前に起こったということを示す概念
c‡complete	目的とした動作の全てを完了するということを示す概念
c‡moment	動作が短時間行われることを示す概念

種々の相を，これらの概念を用いてどのように表すのかを次に示す．

未然相　c‡pred←object－c‡begin
－manner→c‡yet
日本語：　～する
マレー語：aka ＋動詞，動詞＋未来を表す副詞

開始直前相　c‡pred←object－c‡begin
－manner→c‡yet
（将現相）　－manner→c‡soon
日本語：　～するところ，～しようと，～しかけ，～しそう
中国語：　要～了，快～了
マレー語：memulai ＋動詞
英語：　be about to ＋動詞，be going to ＋動詞
仏語：　aller ＋不定詞

開始相　c‡pred←object－c‡begin
日本語：　～し始め，～し始ま，～し出
中国語：　～起来，～上，～開
英語：　begin to＋動詞
仏語：　commencer à ＋不定詞，se mettre à ＋不定詞

開始直後相　c‡pred←object－c‡begin
－manner→c‡already
（－manner→c‡just）
日本語：　～し始めた（ところ／ばかり）
中国語：　剛～起来，剛～上，剛～開
マレー語：sudah ＋memulai ＋動詞
英語：　just began＋動詞
仏語：　venir de commencer à＋不定詞

継続相　c‡pred←object－c‡continue
（進行相）　日本語：　～ている，～中，～し続

け

(持続相)　中国語：在／正＋動詞
マレー語：sedang＋動詞，tengah＋動詞，masih＋動詞
英語：be＋動詞進行形
仏語：現在形，半過去形，etre en train de＋不定詞，ne pas arrêter de＋不定詞

終了直前相　c♯pred←object－c♯end
－manner→c♯yet
－manner→c♯soon
日本語：～し終えるところ，～し終えようと，～し終えかけ，～し終えそう
中国語：快～（完／好／成）了
英語：be going to finish＋動詞進行形
仏語：aller finir de＋不定詞

終了相　c♯pred←object－c♯end
－manner→c♯already
日本語：～した，～してしま
中国語：～了
マレー語：habis ＋動詞，sudah＋動詞，telah＋動詞
英語：have＋動詞完了形
仏語：複合過去形，大過去形，単純過去形，前過去形，前未来形，条件法過去形接続法過去形

完了相　c♯pred←object－c♯complete
日本語：～し終え，～し終わ，～し上げ，～し上が，～し切，～し尽く，～し通，～し抜，～し果た
中国語：～（完／好／成）
マレー語：sudah ＋動詞，telah＋動詞
仏語：finir de＋不定詞

終了直後相　c♯pred←object－c♯end
－manner→c♯already
(方現相)　－manner→c♯just
日本語：～した（ところ／ばかり）
中国語：剛～（完／好／成）
仏語：venir de＋不定詞

状態相　c♯pred←object－c♯state
(結果残存相)　日本語：～している，～してある
中国語：～着
マレー語：masih ＋動詞
英語：現在完了形
仏語：受動態，複合過去形，大過去形

短時相　c♯pred←object－c♯moment
中国語：動詞＋動詞，動詞＋一＋動詞，一＋動詞
仏語：前過去形

視点の表現は焦点となる概念（相情報を表す概念か述語的概念）に対して上記の属性子を付ける形で表現される。

5.意図の表現

話者や主体の意図等に関する情報は，事象や事実をどのような意図や判断でみる，あるいは起こそうとしているのかを表すもので，命令，疑問，断定，推量，願望などムードとして表されていた情報が中心となっている．主にムードや文体で表される情報は，話者や主体の意図、気持ち，判断を表していると考え，その観点から整理した．これらの情報は属性子として表され，次のようなものがある．

(1) 文全体に関する情報

imperative　命令
日本語：～せよ
中国語：原型，～嘛
英語：命令形
仏語：単純未来形，接続法現在形，現在形二人称形，不定法

?　疑問
日本語：～するか
中国語：疑問詞，動詞＋不＋動詞～還是～
英語：疑問形
仏語：?

!　感嘆
中国語：多麼～（呀）!，真～!
英語：!
仏語：!

invite　勧誘
日本語：～しよう

中国語： ～吧(ba)
英語： let us＋動詞
仏語： 現在形一人称複数形，si on ＋半過去形

advise 推薦
日本語： ～したほうがよい
中国語： (還是) ～為好
仏語： il vaut mieux ＋不定詞
faire mieux de＋不定詞

request 依頼
日本語： ～してほしい
中国語： 希望～，請～
仏語： voulez-vous...?, voudriez-vous...?, veuillez..

respect 尊敬
中国語： 貴～，您(nin)

please 丁寧
中国語： 請～
仏語： s'il vous plait, 条件法，接続法

sure 状況からの推量（確信）
日本語： ～するに違いない
中国語： (肯定／応該／応／該) ＋動詞
仏語： devoir, etre sûr de ＋不定詞

maybe 可能性があると思っている推量
日本語： ～するかもしれない
中国語： (会／要) ＋動詞
英語： may
仏語： pouvoir

seem 推察，推測
日本語： ～するだろう，～そうだ
中国語： (可能／要) ～
英語： seem
仏語： sembler, 条件法

duty 義務
日本語： ～しなくてはならない，～するべきだ
中国語： (必須／応該／応当) ～
英語： must, should
仏語： devoir, falloir

grant 許可
日本語： ～してもよい
中国語： (可以／許／準) ～
英語： may
仏語： pouvoir, permettre

need 必要
日本語： ～が必要，～がいる
中国語： 有必要＋動詞
英語： need, necessary
仏語： falloir, avoir besoin de

want 願望
英語： want＋不定詞
仏語： 半過去形，接続法，不定法

will 意志
日本語： ～する
中国語： 要＋動詞
仏語： 単純未来形

rumore 伝聞
日本語： ～するらしい
中国語： 好象～，象是～
仏語： on dit que, 条件法

recommend 評価
日本語： ～するに値する
中国語： (値得／配／可以) ～
仏語： valoir

conclude 断定
日本語： ～したのだ
中国語： 是～
仏語： 直接法

try 試行
日本語： ～してみる
中国語： 動詞＋動詞＋看
仏語： essayer de＋不定詞

voluntary 自発
日本語： ～せずにいられない
中国語： 不＋動詞＋不行
英語： cannot stop ＋動詞進行形
仏語： ne pas s'empêcher de

unwill-duty 不承
日本語： ～せざるをえない
中国語： 不得不～

(2) 文要素に関する情報

emphasis 強調
中国語： 是～
仏語： meme, donc

topic	提題
forcus	焦点

意図の表現で文全体に対する情報は，話者あるいは主体の意図等を表す特別なノード (c#statement)に対して上記の属性子を付ける形で表現される．また，文要素に対する情報は，事象・事実の表現の中の各対象概念に対して属性子を付ける形で表現される．

6.文章構造の表現

文章の構造を表す情報は，文章の構造を中間言語に反映するためのもので，文本体とヘディングの関係や，前後の文との関係に関する情報が中心となる．

文章の構造を表現するために，次のような関係子を導入した．

previous-st	前文
sub-st	従属文
co-st	並列文
quotation	引用文
modify-st	修飾文

文章構造の表現は，話者や主体の意図等を表すノード (c#statement) を上記の関係子で結ぶことによって成される．

7.表現形式

中間言語の表現は，文に含まれるすべての概念間の関係を記述する．中間言語の表現で用いる語彙は，概念見出し，関係子，属性子，および確信度である．記述形式は，タプル表現とグラフ表現の２通りがある．

(1) タプル表現

表現の一般性と記述データの取り扱いの良さを考慮し，関係子は２項関係，属性子は１項関係を表すものとする．タプル表現においては，この関係子と属性子による関係表現を，そのままタプルの形に表現する．すなわち、関係子に関する表現は，

(<関係子>, <概念1>,<概念2>,<確信度>)

であり,<概念1>から <概念2>に向かって <関係子> という関係が <確信度> で規定される性質で存在することを表す．また，属性子に関する表現は，

(<属性子>, <概念>, <確信度>)

であり，<概念> が <確信度> に規定される性質の <属性子> という属性をもつことを表す．<確信度> は，関係の強さを示すもので，1(関係が存在する）と0(関係が存在しない）の２値のみを用いる．

タプル表現中の <概念> は，概念見出しか複合概念のいずれかである．概念見出しは，単語辞書で規定されるものであり，複合概念は，タプルの集合を一つの概念と見なしたものである．この複合概念は，名前を付け，多数の箇所で参照することができる．また，複合概念どうしは完全な入れ子構造となっている．タプル表現の詳細な文法は，次のとうりである．

<概念>::= <概念見出し> | <複合概念>
<複合概念>::= (<関係子>, <概念>, <概念>,
<確信度>)...
(<属性子>, <概念>, <確信度>)
...
<確信度>::= {0,1 }

(2) グラフ表現

タプル表現を理解しやすいように視覚的に見やすいグラフの形に表現したものがグラフ表現である．複合概念は，グラフ全体を１つのノードにすることによって表すことができる．したがって，グラフ表現は，ラベル付き有向ハイパーグラフになる．

8.おわりに

多言語翻訳を目指した中間言語を設計するに当たって，事象の表現，視点の表現，意図の表現，文章構造の表現という４つの観点から文章の表している情報を整理することによって，机上の検討ではあるが日本語，中国語，マレーシア語，英語，フランス語について言葉のもっている概念を除いて，普遍的な表現が得られる枠組みを作成することができたと思われる．今後さらに細部の表現についても検討を加えるとともに，タイ語，インドネシア語についても適用可能であるかの検討を行い，真に普遍的な中間言語の枠組みを作っていきたいと思っている．

本研究は，(財)国際情報化協力センターが通商産業省から委託を受けた研究プロジェクト推進事業「近隣諸国間の機械翻訳システムに関する研究協力」で行った成果である．

参考文献

1)概念辞書 (第１版),日本電子化辞書研究所,
TR-007(1988)