

多言語間翻訳のための中間言語について

石崎 俊 内田 裕士
電子技術総合研究所 (財)国際情報化協力センター

多言語間機械翻訳システムを構築する政府プロジェクトの一環として、電総研、CICCおよび参加近隣諸国と共同で研究開発中の中間言語について述べる。

昭和62年度から6年間の予定で実施している本プロジェクトは、日、英、中国、タイ、マレー、インドネシア各言語を対象とし、多言語間翻訳を高い品質と効率で行うために中間言語方式に近いシステムの構築および基礎研究を目標としている。

中間言語は概念によって記述し、言語依存性の除去を図っている。中でも、本論文では、多言語間で共通性の高い関係概念と属性概念について検討し、それぞれの概念の階層化による分類、定義記述、例文を与えている。

On Interlingua for Multilingual Machine Translation

Shun ISHIZAKI	Hiroshi UCHIDA
Electrotechnical Laboratory	Center of the International
1-1-4 Umezono, Tsukuba,	Cooperation for Computerization
JAPAN 305	3-13-16 Mita, Minato-ku, Tokyo,
	JAPAN 108

ETL, CICC and the neighbouring countries have been developing a multilingual machine translation system supported by the governments. This paper describes contents of interlingua adopted in the project.

It is a six-year project starting from 1987. It treats with Chinese, Thai, Malay and Indonesia as well as Japanese and English. It takes an interlingual type approach to obtain good efficiency and quality for the multilingual machine translation system.

The interlingua is represented with concepts to remove language dependency. Relation concepts and attribute concepts are discussed in this paper which have highly shared properties among the languages. Each of the concepts is classified into hierarchical structure. Definitions and examples are given to the classified ones.

1. はじめに

昭和62年度から始まった「近隣諸国間の機械翻訳システムの研究開発」プロジェクトは、日本語を中心に中国語、タイ語、マレー語、インドネシア語との間の多言語間翻訳システムを6年間で研究開発する。このプロジェクトは通産省のサポートの下で、電子技術総合研究所と(財)国際情報化協力センター(CICC)が中核となり、参加各国と協同で翻訳システムの構築を行うものである。

本プロジェクトでは多言語間翻訳を効率的に行うために、従来のトランスファー方式とは異なり、中間言語方式に近い形式の翻訳システムの研究開発を目指している。いわゆる理想的な中間言語方式では、「すべての国の言語を概念化して共通化し、万国共通の概念を用いて翻訳する」と考えがちであるが、本プロジェクトでは、むしろ「すべての国の言語から抽出される概念の和集合を中間言語のための概念の総体として考え、その中で必要に応じて可能な限り共通化を図る」という方針でプロジェクトを進めている。

本稿では、このような多言語間機械翻訳を研究する上で要となる中間言語について、電総研による文脈情報変換型機械翻訳システムCONTRASTの研究開発[1]と、CICCにおける中間言語ワーキンググループによる調査研究の成果[2]をまとめている。

特に、中間言語の中でも、共通して頻繁に使用する関係概念と属性概念について、それらを階層化して分類し、階層化した各概念の定義と例文をとりまとめて示している。

そのうち、属性概念については、参加各国語において本分類の適合性を検討中であり、各国のnativeの研究者と研究を進めていく予定である。

2. 中間言語のための概念辞書

2.1 概念の定義[3]

「概念」とはインスタンスの集合として考える。ある概念について記述する時、次の4項目に分類して表現する。

- (1) 上位概念
- (2) 下位概念
- (3) 制約(他の概念との関係)
- (4) インスタンスが持つ特徴

上位、下位概念は概念としての包含関係のことで、ISA関係とも言われている。制約は他の概念を用いてその概念を規定するものである。たとえば「運転手」という概念は「乗り物」を「運転する」「動作主」として定義される。4番目のインスタンスの持つ特徴は、概念のすべてのインスタンスに共通のfeatureをまとめて表現しておく。ただし、上位概念に表現されているものは省略して、情報のインヘリタンスが機能するように構造化する。たとえば「カナリア」という概念はカナリアという事例(インスタンス)のすべての集合であり、その共通のfeatureとして、羽の色は黄色、食物は穀類、飛ぶことが可能、などが記述される。

2.2 概念の体系

概念の体系は図1に示すように大きく分けて「物」と「事象」に分類する。「物」は更に生物、人間、コンピュータなどの大量の概念を含む下位階層を持つ。

「事象」概念は「動作」と「状態」に分ける。この分類は文法的なカテゴリとはまったく異なる視点によるものである。たとえば、

「椅子がある。」

「フランス語ができる。」

などの動詞は状態を表わす概念に対応する。

本稿で主として述べる関係概念(いわゆる関係子)と属性概念(いわゆる属性子)は状態概念の下位に属するもので、中間言語を記述する上で最も重要な機能を持つ概念である。

この他に重要な機能を持つものとして、「可能」、「推量」、「義務」などの様相概念があるが、一般的な動作や状態概念として扱い、今後取り入れるために研究を進めていく予定である。

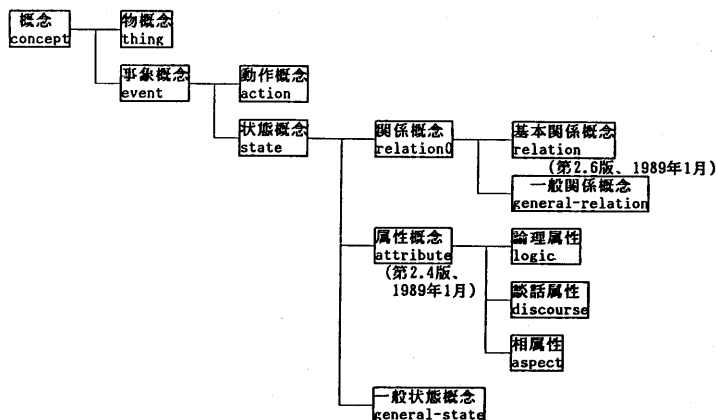


図1 概念体系(上位部分)

3. 関係概念

本章では、いわゆる関係子（スロット）に対応する基本関係概念（図1参照。本章では簡単のために関係概念と呼ぶ）について述べる。従来の関係子の使用法では、図2(a)に示すように、関係子のAGENTやGOALをスロット名として表現し、そのVALUEとして入るものを指す形式をとっていた。しかし、この方法では「AGENTであること」や「GOALであること」を表現することができない（指示できない）。そこで図2(b)に表わすように、AGENTやGOALを概念として定義し、SLOTVALUEやCONTEXTを用いて表現すれば、関係概念を中心とした2引数関係を統一的に表記できる[4]。

従って従来のスロット（関係子）表現（図2(a)）は関係概念による表現（図2(b)）の簡略表現とみなすことができ、支障のない限り使用可能である。

また、図1にある基本関係概念と一般関係概念は大きな相違はなく、前者は中間表現に頻度高く出現する汎用性のあるもので、多言語間でよく比較検討する必要があるものである。

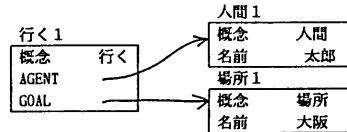


図2(a)「太郎が大阪へ行く」のスロット表現

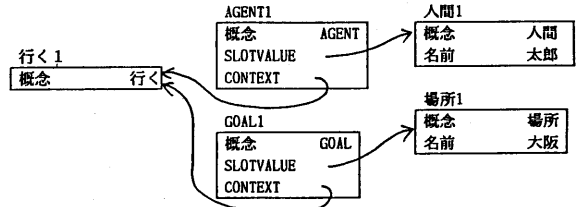


図2(b)関係概念による表現（(a)参照）

3.1 関係概念の階層化[3]

従来の関係子のセットは、階層構造を持たず、フラットな表になっているものが多かったが、次のような理由によって関係概念の階層化を行った。

- (1) 文脈解析では、文章における解析のスコープを広げると、それだけ、二つの概念間の関係が明確化していく。このように、局所的には曖昧な関係概念が解析を進めるに従って明確になるプロセスを、関係概念の階層構造に対応させることを目指す。
- (2) 多言語間で関係概念の統一を図るとき、
 - (a) 言語ごとの関係概念の類似性、詳細性の明確化
 - (b) 関係概念の追加、変更の容易性が可能になる。
- (3) 言語理解・生成技術の進歩に合わせて、関係概念の詳細性（階層構造の深さのレベル）を設定できる。

ただし、翻訳の立場からは詳細な関係ほど良いとは必ずしも言えない。対象とする多言語の間に必要なだけの詳細性を保つようにするべきであろう。

このような観点から、本稿において階層化した関係概念はなるべく多くの種類を列挙し、幅広く多言語間で検討できるようにしている。

表1において実線の四角で囲った関係概念を標準的に使用する目安にしている。また、実際に使用する時はそれより右側の詳細な関係概念を用いることは問題ない。ただし、今後各国言語の例文を増やしていく過程で若干の修正があると思われる。

表1で中央の英文は簡単な定義であり、その右に和文の例文とその英訳がある。

3.2 関係概念の構成要素

(1) 事象内関係

ある事象を中心に、動作主体、対象、時空間対象、対象の変化の始点と終点、時、空間（場所）、利用具・様態・程度などの動作仕様を表わす関係概念からなる。

動作主体は有意志体である動作者と、自動物や現象主を表わす動作物に分ける。ロボットや台風なども動作主体になることができる。

ある動作に複数のAGENTがある場合はそれらを集合として（属性概念を用いて）表わすことを考えている。

時空間対象は、特に時や空間を対象とする事象に使用する。対象の変化の始点および終点は多種類の細分化が可能であるが、それぞれ一つにまとめている。

動作仕様では、道具、手段、方法を一元化して「利用具」を使用する。「程度」や「様態」も含めて、これらの関係概念の利用においては、いわゆる「一文一格」の原則は成り立たない場合が考えられる。

(2) 事象間関係

因果関係を表わすものとして原因・理由、条件、目的がある。その他に、同時、連接(and)、離接(or)がある。内容(content)については、たとえば、

「彼女はそれが素晴らしいと思う」で、
 0 : (agent 思う_1 彼女_1)
 (object 思う_1 @ 1)
 1 : (object 素晴らしい それ_1)

ここで、0 : や 1 : は中間表現の部分構造を示し、@ 1 は 1 : で指示される部分構造全体を指す(topicやfocusによる指示は省略している)[2]。

(3) 物間関係

包含関係として部分、要素、所有、概念包含、例示がある。例示は概念とインスタンスの関係であり、逆の方向の関係を概念(concept)とする。表層の言語表現を概念体系と対応付ける時に必要な関係概念である。

物間関係には、その他多くの関係があると思われるので、今後充実させることとし、それらを物間修飾としてまとめておく。

(4) 物内関係

ある物概念に内在する特徴を示すための関係概念で、属性、単位、数量、値、範囲始点、範囲終点がある。

たとえば、「値段が100円のりんご」の中間言語表現は次のとおり。

(feature りんご_1 値段_1)
 (value 値段_1 円_1)
 (quantity 円_1 100)
 (concept りんご_1 りんご)
 (concept 値段_1 値段)
 (concept 円_1 円)

(5) 状況限定

比較を示すものとして、比較基準、対照、等価・同格がある。対照については、いろいろなレベルにおける対照がありうるので適切な表現構造を用意する必要がある。

また、範囲、役割・視点も状況限定の一種と考えよう。

(6) その他

照応関係は本来の中間表現では、十分な文章解析プロセスを経たあとに必要としないものであるが、一時的なものとして使用できるように用意した。

修飾は以上の関係概念に含まれないその他のもの、どこに含まれるべきか不明のものを扱う。

4. 属性概念

属性概念も関係概念と同様に、多言語間翻訳では重要な機能を担うものである。特に、相属性（アスペクト）のように、言語によって表現方法、形式が異なる場合が多いものについては、その意味の統一、分類がポイントになってくる。

本稿では、そのような属性概念の多言語間における共通化、統一のための資料として、論理属性、談話属性、相属性を分類、階層化して定義と例を与えた。

表2で英文は例文であり、その右に和文で簡単な定義を書いている。

属性の表現形式は、1引数をとるものが多いが、集合属性、否定、疑問などでスコープを伴う表記が必要なものがあるため、各属性概念毎に表現形式を考察する必要がある。

4.1 論理属性

集合に関するもの、否定、定性に分けられる。集合属性では、集合性、全称・部分などの部分性、連言・選言・順序性などの要素間属性、網羅性・排他性などの集合構成がある。

たとえば、「一郎か二郎が大阪へ行った」では

```
(element 集合_1 人間_1)
(name     人間_1 一郎)
(element 集合_1 人間_2)
(name     人間_2 二郎)
(or       集合_1)
(agent   行く_1 集合_1)
(complete 行く_1)
```

となる。(ただしconcept関係は省略)

否定属性は、否定の範囲(スコープ)を記述する必要がある、難しい概念であるが、ここでは概念体系に沿って物、動作、状態に対する否定に分類している。

定性は特定性や確定性とも関連する。ここでは普遍的な真理を表わす恒常性と照応に分類している。照応は、本来表層的な関係を表わすものであり、十分に解析技術が発達すれば必要のないものであろう。しかし、現時点では、指示詞などに対応する照応概念に属性として内部照応または外部照応を付ける。また、内部照応と外部照応を一つにまとめることも考えられる。

4.2 談話属性

談話におけるメタレベルの属性として提題、焦点、強調などを用いる。局所焦点は「りんごが赤い」と「赤いりんご」などの違いを表現するために、局所的に焦点が「赤い」にあるか、「りんご」にあるかを表現するために用いる。

たとえば、「りんごの色が赤い」の中間言語表現は、

```
(concept りんご_1 りんご)
(concept 色_1 色)
(concept 赤い_1 赤い)
(feature りんご_1 色_1)
(value   色_1 赤い_1)
(local-focus 赤い_1)
```

となる。一方、

「赤い色のりんご」では

```
(local-focus りんご_1)
```

となっている。

参考のために概念体系における分類は以下のようになる。

```
(is-a 赤い 状態)
(is-a 色 物)
(is-a りんご 物)
```

仮定法、命令法については、様相概念として扱う可能性があり、今後検討する予定である。

疑問は否定と同様にスコープを伴うものであり、その表現形式を検討する必要がある。

敬語については、日本語に基づいて、丁寧、尊敬、謙譲に分類した。たとえば「申す」という動詞は専ら謙譲に用いられる。しかし、表2のその右側に記述しておいたように、敬語は談話解析の一つの解析目標であり、将来十分に解析可能になれば、このような細分類は除かれるよう。

4.3 相属性

相(アスペクト)は時制(テンス)と完全に直交した概念ではなく、相対的にからみあったものとして考えた方がよい[5]。国語学でも明確な定義や分類が確立している訳ではないが、本稿では英語も含めた多言語間で共通に使用できる相属性の分類と定義を試みる。

まずテンスであるが、これはアスペクトではなく時概念とし、過去、現在、未来に分類する。これらの時概念は時間座標軸の基準点(defaultはその文の発話時点)からみたときの時の区分である。関係概念の一つの「時 time」は正確には「時間関係」のことなので区別する。また、常に成り立つ「超時」は論理属性の「恒常性」に含める。

アスペクトは動作そのものに内在したり、助動詞や副詞を用いて表現する。また、図1に示すように、「動作」と「状態」は上位の概念レベルで分離しているので、動作アスペクトと状態アスペクトの区別はここでは行わない。

終了相は完了相、結果相、直後相を含む。完了と結果の違いは、動作が終了したことに着目するか、終了した結果の状態に着目するかの違いである。

特に、中国語では結果動詞という二つの部分からなる合成動詞があり、第二の部分で結果を示す形態をとる[6]。日本語でも以下に掲げるような継続動詞(概念)と瞬間動詞のそれぞれに結果動詞と非結果動詞の区別を組み合わせるという説もある[5]。

開始相は開始相と直前相に分かれる。開始相は「開始する」という動作概念に関連するが、中国語の「起来」によるアスペクトとして挙げている。

反復相は動作の反復と状態の反復に分かれ、ある時間間隔をおいて繰り返すことを表わす。

継続相は、動作が継続して起こることで反復とは区別する。また進行相は継続相に含めて考えている。未完了相は現在に至るまで動作や状態が継続し、まだ、完了していないことである。

瞬間相は動作が瞬間的に生じることで、「ひらめく」や「出る」など動作そのものが持つアスペクトである。

この他に、「経験」については様相概念に含める予定である。

5. 終わりに

中間言語の中でも中核的機能を持つ関係概念と属性概念について述べた。

本文で示した分類や定義は、現時点では「近隣諸国間の機械翻訳システム開発」プロジェクトで研究を進めるための基礎であり、多言語間翻訳に使用するために、日、英、中、タイ、マレー、インドネシア各国語に基づいた検討を行なっているが、今後も例文を増やし精度を上げる予定である。

中間言語表現のためのシンタクス[2]については別に詳しく発表する機会を得て明らかにしたい。

また、複合語の処理については、複合語の内部構造の解析を進めて効率の良い辞書作りを目指すことを考えている。

最後に、本稿における関係概念および属性概念は必ずしもすべてが現在の自然言語処理で解析・生成できるものばかりでない。近い将来の当該技術の進歩を予想して、中間言語を表現する枠組みを用意し、一つの指針とする考えである。

謝辞

本研究に御協力いただいた電総研中島隆之知能情報部長はじめ自然言語研究室の皆さん、CICC辻専務をはじめとする機械翻訳研究所の皆さん、CICC技術委員会、中間言語ワーキンググループの皆さんに感謝する。

参考文献

- [1] Ishizaki, S., Sakamoto, Y., Ikeda, T., and H. Isahara, "Machine Translation Systems Developed by the ETL", Future Computing Systems, Oxford University Press, Vol.2, No.3, 1987.
- [2] CICC中間言語ワーキンググループ、関係子および属性子資料 1986-1988.
- [3] 石崎、井佐原、橋田、内田、横山、「文脈理解のための概念記述法」情報処理学会自然言語処理研究会 64-7 1987.
- [4] Hasida, K., Ishizaki, S. and Isahara, H., "An Approach to Abstract Generation," Bulletin of ETL, Vol.52 No.4, pp.551-564, 1988.
- [5] 金田一春彦、「日本語動詞のアスペクト」むぎ書房 1976.
- [6] ヤーホントフ著、橋本訳、「中国語動詞の研究」白帝社 1987.
- [7] Ishizaki, S., Isahara, H. and Handa, K., "Natural Language Processing System with Deductive Learning Mechanism," in Nagao ed., "Language and AI," Elsevier Science Publishers pp.125-144, 1987.

[8] フィルモア、田中春美訳「格文法の原理」三省堂 1975.

[9] 日-英科学技術用語研究データベースの開発に関する報告書 1986.3.

[10] 樋口、戸田「会話文処理情報格構造」in「認知理論に基づく社会的相互作用過程の解明」1987.9.

[11] Dong, D. Z. and Deling, Z., "KY-I MT System and Some Linguistic Aspect", Proc. of MT Meeting in Peking 1987.

[12] 河部利夫、「基礎タイ語」大学書林 1988.

表 1 基本関係概念(第 2.6 版)

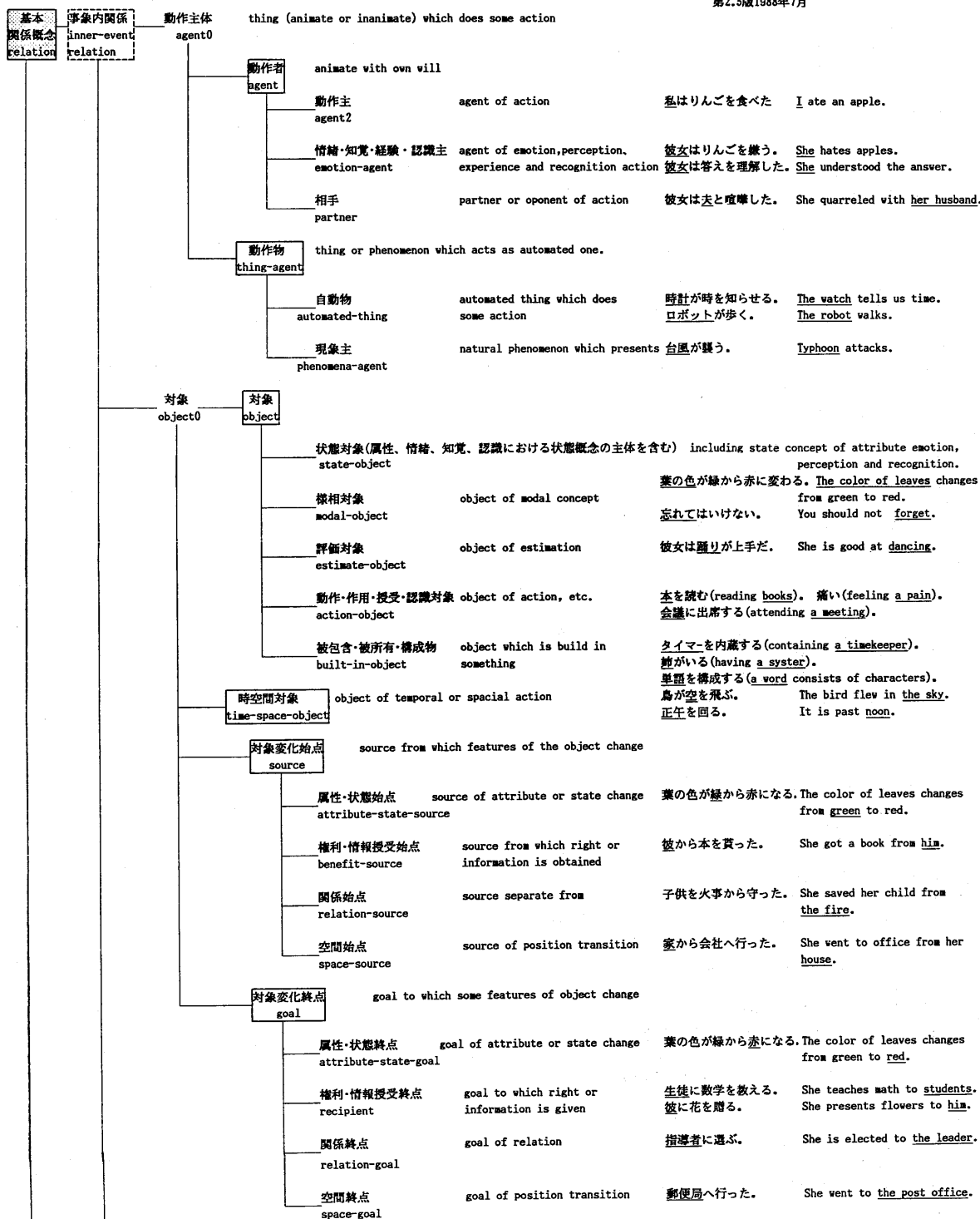
Hierarchy of Relations(the 2.6th version)

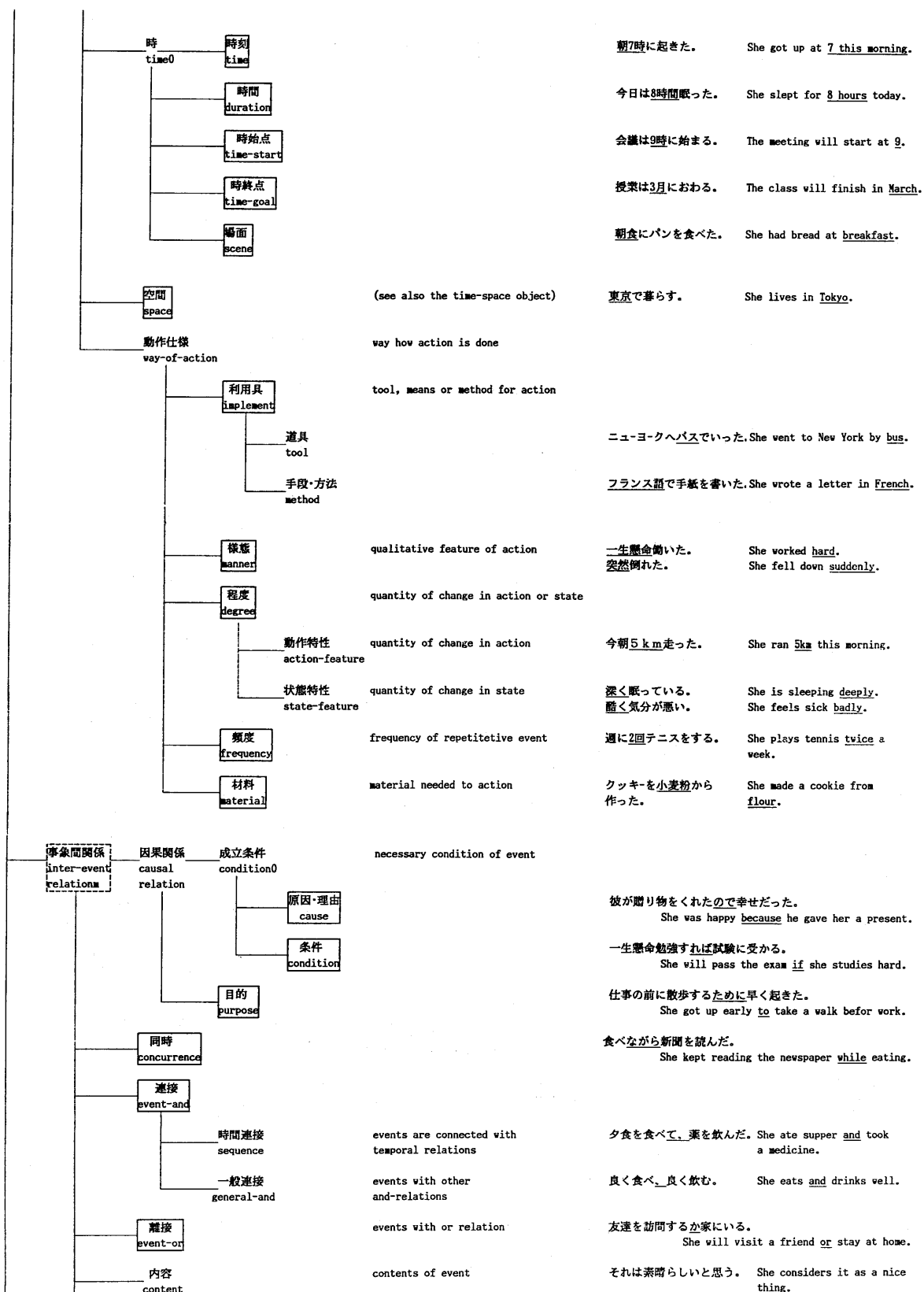
第0版1986年12月

第1版1988年 1月

第2版1988年 3月

第2.5版1988年7月





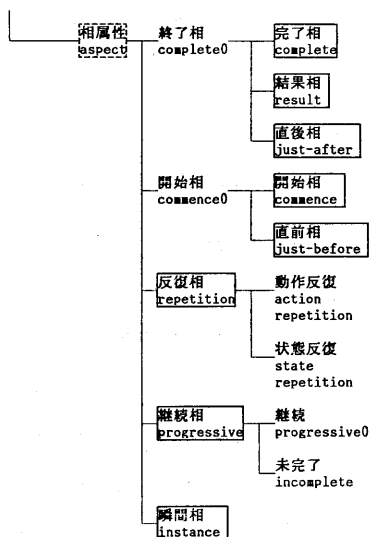
物間関係 thing-relation	包含 inclusion	部分 part	part-of relation between things	自動車のドア。 <u>doors of</u> the car.
		要素 element		委員会のメンバー。 その家族の娘。 <u>a member of the committee.</u> <u>a girl of the family.</u>
		所有 owner		花子の犬。 <u>Hanako's</u> dog.
		概念包含 is-a	is-a relation between concepts	動物のカメを買った。She bought a turtle <u>of an animal.</u>
		例示 instance	instance-concept relation	(その犬がそれを食べて死んだ。 <u>The dog</u> ate it and died.)
	物間修飾 modifier		miscellaneous relation between things	
物内関係 inner-thing relation	属性 feature		feature contained in a thing	りんごの数 <u>number of</u> the apples. その花の色は赤だ。 <u>The color of</u> the flower is red.
	単位 unit		unit attached to an amount	5個で300円。 300 yen <u>per 5 pieces.</u>
	数量 quantity		relation between a numeral and an amount	5個の林檎を食べた。 She ate <u>5</u> apples.
	値 value		relation between an attribute and its value (value is not restricted to numerals)	その花の色は赤だ。 <u>The color of</u> the flower <u>is red.</u>
	範囲始点 from		starting point of thing attribute	2ヵ月から5ヵ月までの休暇。 vacation from 2 months to five months.
	範囲終点 to		final point of thing attribute	1から10までの数字。 digits from one to ten.
状況限定 specification	比較 comparison	比較基準 comparison0	criteria for comparison	母親よりも背が高い。 She is taller than <u>her mother.</u>
		対照 contrast	contrastive relation among concepts	ジョンは東京にいるが、妻は京都にいる。 <u>John is in Tokyo, but his wife is in kyoto.</u>
		等価・同格 equivalence	equivalent relation	東京、日本の首都。 <u>Tokyo, the capital city.</u>
	範囲 range		range to specify the situation	クラスで一番奇麗だ。 She is <u>most beautiful</u> in <u>her class.</u>
	役割・視点 role		role or view in the situation	
		役割 role		アドバイザーとしてその方法を推薦する。 She recommended the method as an <u>advisor.</u>
		視点 view		年齢の割りに若い。 She looks young for her <u>age.</u>
	照応関係 reference		relation among anaphor and antecedent (This is used temporarily until anaphora analysis is implemented, see "endophoric")	探偵小説を昨日買った。一日でそれを読んだ。 She bought <u>a detective story</u> yesterday. She read <u>it</u> in a day.
	修飾 general modifier		ambiguous relations not specified by above ones or miscellaneous relations not included above	

表 2 属性概念 (第 2.4 版)

attributes

第0版1986年12月
第1版1988年 1月
第2版1988年10月

属性概念 attribute	論理属性 logic	集合属性 set0	集合性 set		
			部分性 some0	全称 all	people, audience 集合を表す概念
				各々 each	集合に含まれる全ての要素
				部分 some	集合の各々の要素を意識しながら 集合全体に言及
			要素間属性 element attribute	連言 and	集合の一部分 (nullは含まない)
				選言 or	挙げられた要素のすべて
				択一 either	挙げられた要素のうちの一つまたは複数
				順序性 ordinal	挙げられた要素のうちの一つのみ 選択
			集合構成 set con- struction	網羅性 exhaustive	集合中の要素間に順序がある
				排他性 exclusive	ある集合が下位概念ですべておこ われること
				等 so-on	複数の集合が互いに共通部分を持 たないこと
					その他にも要素があること
	否定 negation	事象否定 not-event	動作否定 not-action	She does <u>not</u> go.	動作概念の否定
			状態否定 non-state	He <u>is not</u> free.	状態概念の否定
		物否定 non-thing		nonfiction, nonsmoker	物概念の否定
	定性 definite	恒常性 truth		The sun rises <u>every morning</u> .	普遍的な真理
		内部照応 endophoric	前方照応 anaphoric	John went to the post office because <u>he</u> must sent a letter to his wife.	照応概念(he)が前出の先行概念(John)をうける場合
			後方照応 cataphoric	When <u>he</u> arrived at school, <u>John</u> was too late.	照応概念(he)が前出の先行概念(John)をうける場合
		外部照応 exophoric		Take <u>the</u> bag in front of you to your mother.	言語表現の外にある実体の指示 (see "reference" in Relations)
談話属性 discourse	話題 topic			The <u>nose</u> of elephant is long. <u>Japanese</u> is difficult to learn.	談話中で話題化された概念
	焦点 focus			Elephants <u>have long noses</u> . He went to <u>Kyoto</u> , not Osaka.	談話中で焦点となっている概念
	局所焦点 local-focus			The apple <u>is red</u> . A red <u>apple</u> .	談話の中で局所的に焦点となっ ている概念
	強調 emphasis			It is <u>he</u> that broke the window.	談話中で特に強調された概念
	法 mood	仮定法 subjunctive		If I <u>were</u> rich, I could help you.	反事実的仮定 (関係概念「条件」 と区別する)
		命令法 imperative		<u>Work hard</u> or you will be fired.	動作・状態概念に関する命令
		疑問 interrogative		<u>Do</u> you go to Kyoto today?	疑問 (あるスコープを伴う)
		敬語法 honorific	丁寧 polite	<u>Would you mind</u> helping my study? I am <u>hoping</u> you will give me some advice.	[敬語に関する基本的な構造は話者と聞き手の相対的な関係と、話者が敬語を使用するときに基準とする位置に依存する。このような構造が解析できるまでは左のような分類を用いる。]
			尊敬 respect	There are no <u>honorable</u> changes.	
			謙譲 modesty	I accept your apologies <u>with humility</u> .	



He has read the book.

ある動作が終了し、その終わったことに焦点がある。

He got married last year.

ある動作が終了し、その結果としての状態に焦点がある。

The train left just this moment.

ある動作が終了し、その直後に焦点がある。

The baby burst into tears.

ある動作が開始し、その始まったことに焦点がある。

The girl is about to cry.

ある動作が開始する直前に焦点がある。

He reads a newspaper every morning.

動作がある間隔で反復する。

She is wearing glasses whenever it gets dark.

状態がある間隔で反復する。

He is running.

動作が継続して生じていること。

We have lived here all our lives.

動作・状態が現在に至るまで継続していること。

The nice idea frushed on him last night.

ある動作が瞬間的に起こること。