

## 前置詞分析のための基本的手順について

神 博史      鈴木 雅美      野垣内 宏      若菜 弘志  
 (KDD研究所)                                      (日本IR)

### 《はじめに》

この報告では、KATEシステムの翻訳対象である、ORANGE BOOK<sup>(1)</sup>に出現する前置詞の諸相に着目しながら、翻訳用前置詞辞書作成のための前置詞辞一タについての分析手法・手順について述べる。

辞書辞一タの作成において、動詞や名詞、形容詞等のいりやりの内容語の分析に比べて、前置詞の機械語の処理はより不十分なものがあるのが一般的である。私には、これらの諸相と前置詞との関連情報（たとえば、結合情報）を考慮しつつ、前置詞の側からの分析の辞書辞一タの作成を志す、かつできるだけこれを定式化して記述しようと試みている。この報告はこうした試みの中間的な作業報告であるが、前置詞の用途的機能を考察すること、動詞等の内容語の分析辞一タを模範とし、より精密に分析することも可能になると思われる。

### 《KATEシステムについて》

KATEシステムは、電信電話通信に際する動詞文書、通称 ORANGE BOOK（総単語数約1万語、単語語数約2,500語、総文例数約2,500文）を翻訳対象とする英→日機械翻訳システムである。現在は構文解析レベルの実験を主に行っている。パーサーには拡張リジビを修正して用い、解析用の文法規則は約600個（適用制限条件付き）、解析用単語辞書は約5,000項目程度の規模である。解析段階では、単語辞書、語彙表規則のほか、禁止木、排他木データを用い、トーチメント方式による選択法で、唯一の解析木を出力するものである<sup>(2)</sup>。

解析用辞書項目の一般的内容は、動詞項目（ホーンビーコード<sup>(3)</sup>（細分類コードを含む）、結合情報（特に前置詞との）、シリーラス情報（体系的にはない）<sup>(4)</sup> などを含み、名詞項目（可算・不可算情報、結合情報、シリーラス情報<sup>(5)</sup>）、形容詞項目（KATEコード<sup>(6)</sup>（総語情報をベースに独自で作成）、結合情報、シリーラス情報<sup>(7)</sup>）を含むものである。

テキストの ORANGE BOOK は、1文あたりの平均語数の語で、電信電話通信分野に特有の語彙傾向をもっているが、分析対象語の前置詞として下表1のようになっている（辞前置詞<sup>(8)</sup>を除く）あり、総語数に対する前置詞全体の（辞前置

『表1- ORANGE BOOK 前置詞リスト(頻度順)』

- |                   |                       |                    |                    |
|-------------------|-----------------------|--------------------|--------------------|
| 1. of (2570)      | 14. per (96)          | 27. below (12)     | 35. excluding,     |
| 2. in (1287)      | 15. during (81)       | 28. before (10)    | toward, versus (2) |
| 3. to (1071)      | 16. within (75)       | 29. plus, through  | 36. beyond, down,  |
| 4. for (1041)     | 17. into (69)         | (9)                | regarding,         |
| 5. at (518)       | 18. under (62)        | 30. concerning,    | throughout (1)     |
| 6. on (491)       | 19. without (42)      | considering (8)    |                    |
| 7. with (479)     | 20. after (30)        | 31. until (7)      |                    |
| 8. by (350)       | 21. including (26)    | 32. among (5)      |                    |
| 9. as (275)       | 22. against (25)      | 33. along, inside, | ** ( ) は使用         |
| 10. from (254)    | 23. except (23)       | minus (4)          | 回数                 |
| 11. between (244) | 24. above (17)        | 34. about, outside |                    |
| 12. than (125)    | 25. via (16)          | (3)                |                    |
| 13. over (104)    | 26. across, upon (14) |                    |                    |

詞を含む) の割合(約14%)である

《作業方針について》

冒頭で述べたように、私たちが、前置詞の側から文構造を見ようとするのだから、翻訳システムにおいて、対象とする文例の解析と程度の別差は、文構造の中枢を占める述部を軸に対立するものであるが(動詞と形容詞の統語構造の解析と格構造の把握のように)、文構造の中心から周辺に広がる範囲にあって、様々な関係概念を表出する前置詞句の分析は、翻訳システムの質をより向上させるうえに極めて重要な作業であるといえる。

そこで、前置詞分析に関わる3-2をおおむね関連する2つについて整理すると、おおよそ以下のようになる。

- (I) 用法・意味(深)の分析等 → 前置詞分析表 (表2参照)  
(多数格解消の仕組みの分析等も含む)
- (II) あいまい性解消のための分析等  
\* システム検証用ノート作成  
シリーフス用  
ノート作成  
(表4参照)
- (III) 分析不可能等への対処に関する検討

ここで、(I)の項目は、通常の辞書情報の基礎となる等々の研究であり、このように、前置詞の統語環境および意味環境、さらに訂正日本語の選択条件などを明らかにするものである。すなわち、ここで前置詞のVxシカレ等々(A + preposition + B)における; Aの形態-品詞はVと意味素性、Bの形態と意味素性、AとBの関係概念-場所、時間、原因、目的等、AとBの結合の態-強、弱、位置等、訂正日本語、前置詞コード等)を個別にシステム上の等々を扱ひ、分類的に一般化して記述するのである(表2参照)。

このようにして、(II)の項目(すなわち本文の構造に関する結合関係の問題を、前置詞を中心に整理する)の方向に早急な研究が必要である、という方向性から検討される等々の作成である。これとは、

A) The principal characteristics recommended for a modem to transmit data at medium speed in the general switched telephone network are as follows :

(一般に採電路網において中速度データを伝送するために用いられるべき主要な特性(状況の通り)である)

B) ...the transmitting sets use the signalling frequencies of push-button telephone sets to transmit data to a central receiving set via the switched telephone network.

(伝送装置が、採電路網を通して中央受信装置にデータを伝送するために(パルス式電話の信号周波数を使用する).....)

C) The access is gained to the responding equipment.....

( 対応装置へのアクセスが実現される ..... )

いづれにせよ *with regard to* の前置詞 (句) が前節の *at* の語句と結びつくわけは、経路論的、意味論的あるいは形式的な観点から一般的に考えられている事柄である。狭い範囲でしか言及されない *with regard to* の場合、(A) のように、*transmit* と *at* と *in* に身をおいた前置詞を結びつける種種的理由 ( *with regard to* ) を、*data* との関係を同様と見做す。しかし、このようにして *with regard to* の前置詞 (句) と先行する主節語句との比較に、この前置詞の判別が可能とされるようでは前置詞側の分析が不可欠であろう。この (B) の作業は *with regard to* の同義的表現を一つ分析してはくまり構造的な分析作業と見るのである。このための前提として私たちは必要としている (B) のレベルの問題を *with regard to* の前置詞の入り方のあいまいさとして頻繁に指摘される構造—実際のテキストの中でどう処理されているか—についての全体的な分析が不可欠である。この点の章では、この問題の分析作業について報告する。

踏まえて *with regard to* (B) の分析は、これは、(C) と (D) の分析に、これも対処不可能な場合、これをどう扱うかであるが、単純に切り離して個別の文を対象とする限り、この種の文構造が存在するのみであるが、一応、単純に処理できるかどうか、対象テキストの背景知識が活用できるかどうかの点の整理が必要であろう。

### 《作業マニュアル》

前置詞の語相を知らぬための分析を針としてあげた3つの項目中、機械翻訳システム作成に必要となる項目は「あいまいさ」に属する項目 (B) について考えることにする。「あいまいさ」という概念について、意味論的観点からの考察を、構造的な観点から *with regard to* の文法を、さらに、前置詞 (句) の入り方の実際の分析を *with regard to* の分析と見做す。

この作業の参考になる書籍は、ORANGE BOOK の文例中、前置詞に *with regard to* の部分から他の部分と関係する部分、この語法明確であるのか、あるいはその程度不明瞭であるのか、を実際に検証することに、テキストを

『表2-前置詞分析表 (部分)』

ABOUT

主要語	前置詞	目的語	CASE	コード	対応日本語	品詞	用例	備考
NOUN +abs, anim. 前置詞と名詞活動 に接する物事	with regard to; on the subject of 主題・焦点	NOUN ±abs	REFR	FP1A 主	N <sub>0</sub> の同様の N <sub>g</sub>	P + V	a book about the stars/ information about special noise sources (OR) + NOUN-gossip, rumor, story, fact; instruction, article, opinion, advice..	O.B. 文中 「about」の用法 on, of, with regard to, concerning, in
NOUN -anim	surrounding 場所の周囲	NOUN -anim	LOCAT	FP1A -	N <sub>0</sub> の同様の N <sub>g</sub>	P + V P	the high wall about the prison/ a fence about the play ground	around, round の用
VERB 動詞活動 名詞活動	with regard to; on the subject of 主題・焦点	NOUN ± abs	REFR	FP3A 主	N <sub>0</sub> の <i>with</i> ～句	P	We talked about our plans. + VERB- chat, joke; debate, comment, consult, speak, say, ...	talk about, vph 動詞. about which ... talk about speak about + vph. which, about which—speak about

体の難易度もまたある程度予測することができると思われる。

そこで、多くの実例を人手で検証するうえで、より効果的作業のためのアルゴリズムを設計し、これに基づいて分析を行おうと、和訳には漢語の表裏にあらうというフローチャートを作成した。

### 〔フローチャートの作成について〕

このフローチャートは実際のテキスト中で前置詞(句)がどのような位置にあり、どのように出現するかを想定して作成したものである。前置の語句との係り方を判別する作業を一回ねら、前置にある語句に、その係り方が容易に識別できるように配慮して作成した。

テキスト中の前置詞の前方にほとんどかの語がある場合、それは以下のよう増品詞の語である。

名詞 / 代名詞 (疑問代名詞含む) / 助詞 / be 助詞 / 形容詞 /  
副詞 (副詞的、辞含む) / 前置詞 / 接続詞

この以外に、コンマ (前置の挿入句がある場合など) やパーレン (前置詞句がカマコクに入っている場合) などもくまが、このフローでは除外してある。また、文頭に前置詞がくる場合も、当然前置詞の前方に語句(代名詞)に付く。

### 〔検証例について〕

これを、次のように文について試してみよう。

a telephone channel not simultaneous with data transmission in the forward direction....

この文は、前置詞 with の前方に形容詞の simultaneous があることで、この語と with の結合関係のチェック (simultaneous の辞書項目を参照する。 simultaneous == ADJ \* ((adjcate ap01 ap05 ap07 ap08) WITH) ) によって、この語と with が強く結びつくことがわかる。

( I was happy in my boyhood. この文は、in と happy が強い結合関係(代名詞)に、in が was happy という動詞句に係ることを判別する。

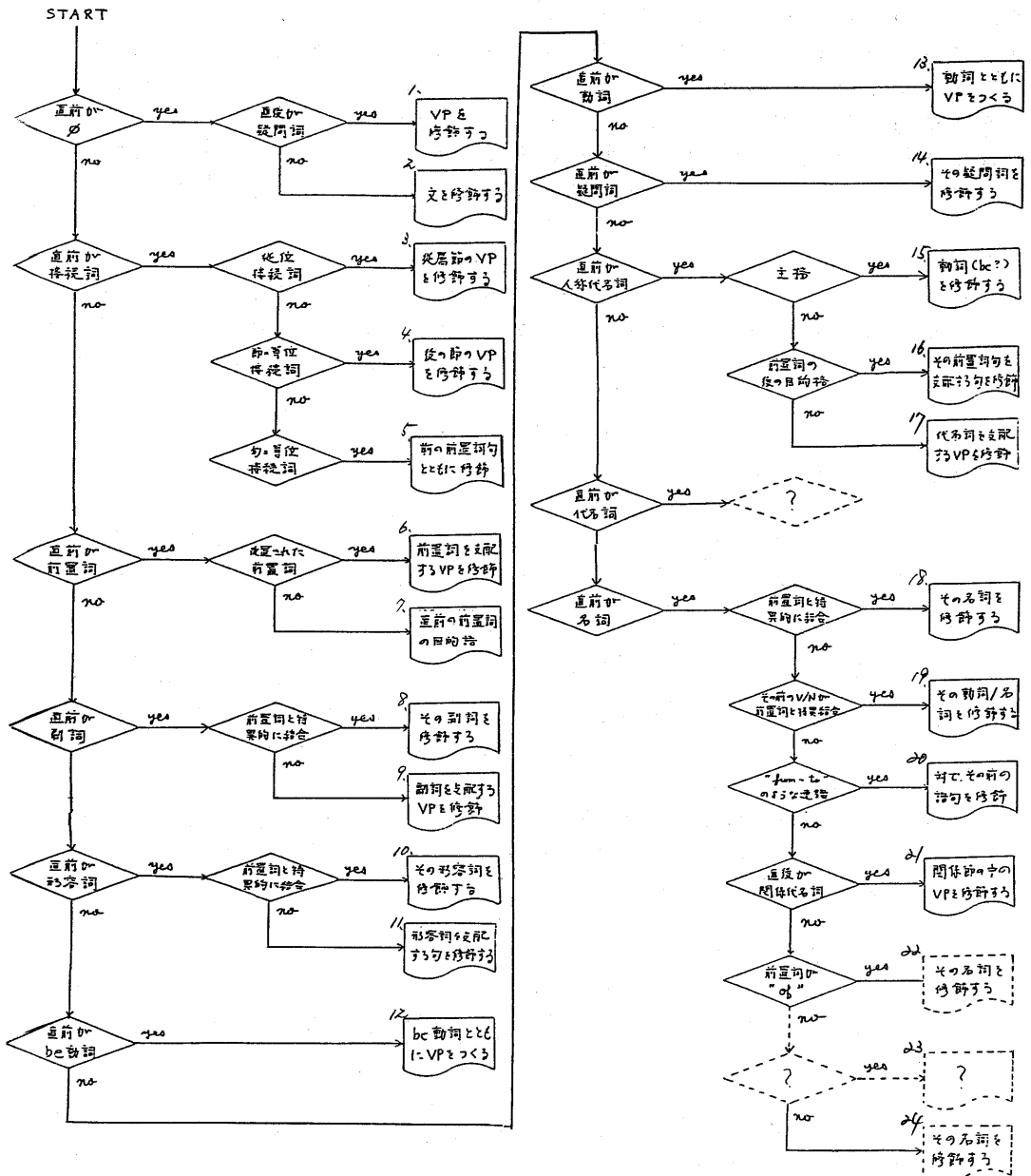
It is recommended that for synchronous transmission the data signalling rates should be divided into two distinct classes ..

この文は、that が接続詞 (従属節を導く) であることから、後続の節の主要述語を修飾する前置詞(句)と判別する。

( that の、ほかにも形容詞、代名詞などがあるが、 )  
( It ... that ~ の形から優先的に接続詞と判別する )

この文のもう一つの前置詞 into は、先の simultaneous と同様、divide と強い結合関係があることがチェックされる ( divide == VP14 \* (VO VP14 (valenc by into between) ) ) 。

# 『表3-フローチャート』



\*\* 前置の語以外に前置詞句用法は 70-a, 7, 8, 10, 12, 13, 14, 18, 22, 23 以外不可  
 \*\* さらに「文修飾」以外に、「述節修飾」とも呼ばれるが、その別名はこの70-aの範囲外である。

It is desirable that the transfer of timing information across the interface shall not be ..

だが、informationとacrossの間に強い結びつきがあることは確かであるが、それ以外では、transferとの間に強い関係を認めることはできないが、それを比較すると、transferとacrossとvia, through, over etc. 場所(経路)を越える際の結びつきは、その間に渡り越えることによる(transferの決意の結合情報項目中の、「決意の結びつき」の扱いが情報) 、この場合acrossはtransferに比べて判定可能。

{ 検証結果について }

環の70-キートを用いて作成した検証結果(以下の通り)である。

\*\* 検証に使用した資料はORANGE BOOK中の200文、以前の文の1/4に作り直した70-キート( S.A. )中の論文中の100文(従来のように)。

	ORANGE BOOK	S.A
直前の語を修飾するもの	88%	87%
直前ではないが直前のものより修飾するもの	10%	13%

\*\* ORANGE BOOKの例文(この70-キート)のうち2%あり、

\*\* 直前の語を修飾するものの中(意味的でないものは除く)の割合はORANGE BOOKは20%、S.A.は12%程度であった。

\*\* ORANGE BOOKと同様の70-キート、直前YELLOW BOOK<sup>(5)</sup>中の100文について検証した結果、直前の語を修飾するもの85%、直前以外の語を修飾するもの11%であったが、この70-キートでは直前の語を修飾するものが4%あり、

この検証より前置詞の採り方の一般的な傾向が明らかとなり、同時に意味的でないもののあいまいさが直前前置詞全体の16%程度あり、また、70-キート中の項目ごとの場合、

前置詞は直前の名詞を修飾 / 直前の動詞を修飾 / 直前の名詞を修飾 (強い結合関係がある場合) / 名詞を修飾 / 直前以外の名詞と名詞を修飾する / 直前の名詞を修飾 (強い結合関係がある場合) / 直前の動詞を修飾する .....

のうちの割合順と比べると一致する。実際の70-キート処理は、このように前置詞の採り方の不一致、解析プロセスの不一致も利用を以てしてあるが、相互の前置詞ごとの不一致は、解析結果を以てして、これを後者情報に記述することにより、その解析結果を以てして設定して

予て述べた通りである。

\*\* 次に、of の二相をとりかえ (上記の文例を)、その 97% が  
直前の後を補脚してあり、この二相的的分弁一々を別開可能な二相  
解析結果を向上させることも可能である。

\*\* あいまい性を排除して読者のため、結合情報等のレクンカルシデ  
ンク処理を怠らないもので、この解消のためにはテキストもそれ  
の二相的的分弁が必要である。

.. that arrangements should be made when necessary to reverse the direction of  
the parity unit at the input and output ....

### 《より精緻な分析のための視点》

検証のためのフローチャートを作成し実験を行ったに、レ、テキスト中の前  
置詞の諸相の分析結果は明白に示されることができたが、この中で「あいま  
いさ」のその消滅への二相的分解を課題として設定した。しかし、私たちが  
のテキストに示すものの比率は、前置詞用語全体の 16% 程度であることがわ  
かるとは十分な収穫である。

次に、「あいまいさ」のその消滅のため、私たちがより分析的な作業を  
想定し、前置詞諸相情報としてこれを活用してみたいと考えている。次に、前  
置詞に関する用語法を整理すると、以下のようになります。

### (I) 前置詞と先行名詞の反配力が強いもの

These theoretic values do not depend on the modulation rate.

This is subject to a mutual agreement between Administrations concerned.

\*\* このうち、depend on と subject to は、前置詞の個別の意味を想  
定するのは十分可能でなく (前置詞の単位の指標として機能する)、前  
置詞諸相情報として整理しにくい例である。これら二相の前置詞  
と先行名詞の結合情報として記述されている。

\*\* この範囲では先行名詞の反配力について、統一的な場合と、意味  
的同意 (agreement between a case) のみならず含むことができる。

### (II) 前置詞と先行名詞の反配力が決定的に強いが、意味的に反配 力を想定できるもの

...are specially well suited to the transmission of signals by a modulator...

\*\* 名詞 transmission (動詞 transmit) は派生したもので、その格  
構造 [Agent, Object, Goal, (Instrument)] をその子系採得  
してこれを伝送する、by (手段格) として transmission の伝送対象  
としての (signal) (対象格)。

(Ⅳ) 前置詞と先行有語句との関係は、一般的に語彙情報也、テキスト  
 不十分の抽出と小規模情報中の判定は乏しいもの  
 \* 同様に表制に於いては不語彙情報及び要求を以て思ふべき

...other international in the general switched telephone network may still support operations at 2400 bit per seconds ....

It is useless to seek the meaning or utility of a book between its covers....

このように、分析の範囲を分ける場合、和文では通常「あいまいさ」を以て問題視してゐる構造レベルのものがあるが、是れが思へてゐる(すなわち)である。

先づ「with」の「with」のレベルは、16%程度のものが「あいまいさ」を以て語彙情報と見做すことができるが、(Ⅱ)と(Ⅲ)のレベルに属してゐるものがある。しかし、テキストの知識も、文脈上の意味の減少を以ての情報の減少(すなわち、同様の語彙情報の担う範囲が狭い)である(Ⅳ)のレベルの構造について(Ⅱ)のレベルに除外して扱ふことが必要である。和文の抽出から見れば、(Ⅱ)のレベルの「あいまいさ」は前置詞用法の5~10%程度である。そのうちが初級レベルの「あいまいさ」のうち7割の割合で、(Ⅱ)のレベルのものである。すなわち、語彙的語彙情報の精密化を以て、解消されると思ふべき。尤もこの実験を以て、between と with の分析を以てした。

[with と between の間]

	with	between
(Ⅰ)のレベルのもの	36%	10%
(Ⅱ)のレベルのもの	53%	29%
(Ⅲ)のレベルのもの	11%	11%

\* with の使用回数テキスト全体に 109回, between に 22回。  
 \* 同前置詞は除外した

この結果から、和文(すなわち)のテキストにおける前置詞(句)の類似的構造について90%程度の(Ⅱ)と(Ⅲ)のレベルの情報を以て、和文の語彙的語彙情報から解析してと判断してゐる。その(Ⅱ)および(Ⅲ)の分析を以てして、前置詞と関係有語句の意味的類似性をより明確に判定する資料を集めることが可能であると、これらから、with と between の先行語の中

collaboration, relation, correlation, relationship, cooperation \*

interface, interaction, interchange, 及び distance, difference, division, imbalance

etc. の語彙情報, 「関係」的の概念(スプレッドの要素が存在しては行を指示)のときもこれらからとることができ、同様に agree, connect, distinguish, divide,



differentiate, distribute などの前置詞や副詞, 動詞といふ語群について, この「関係」概念を捉えることができ, 大抵の語の意味的付着過程 (シソーラスの基礎といわれる) をつかみ取ることもできるわけである。(後年参照)

前置詞を軸としてこの語の語彙的情報の整理がなされ, 先行の (Ⅱ) のレベルの分析から, 一般的の意味のあいまいな構造のかわりの部分についてより具体的な解析を行うことができたことがわかった。その (Ⅲ) のレベルに踏まえてこの構造の他の部分について, 構造的知識の整理がなされたこととし, (Ⅳ) のレベルの分析をふまえてこの解析の成果を予測している。

《おわりに》

関係概念の表裏をよく担う構造から前置詞(句)について, 実際のテキストの分析を通して, その意味と用法の諸相を調べたため, 効果的のフローチャートによる検証, および語彙データの精査化などの作業を行った。このうち, 従来の前置詞(句)に関する「あいまいさ」を解消するにむくむくとのうちのかわりの部分が解消され, より具体的な解析結果を得ることができた。今後にはAIEEなどの解析用データとして活用可能なものの体系的な記述として定式化していくことを目指している。

《謝辞》 日頃御指導頂くKDD研究所鍛冶所長, 野坂副所長, 博松次長, 武田第一研究所室長に感謝する。  
前置詞分析作業に担当した, (株)日本IRの斎藤英夫, 石川に, ソフト開発担当の, (株)SCCの田中氏に感謝する。

文献

- (1) CCITT SIXTH PLENARY ASSEMBLY ORANGE BOOK VOLUME VII.1 1976
- (2) 電信研指論機構研究室, 拡張LINGOL, 1978
- (3) 柳, 橋本, 鈴木, 野垣内; ファルタリング操作を伴うページングの一形式; 自然言語処理研究会資料A3, 1984年5月
- (4) SCIENTIFIC AMERICAN Sep. 1972 Volume 227; Communication by J.R. Pierce
- (5) CCITT NINTH PLENARY ASSEMBLY YELLOW BOOK 1980

〔前置詞関係〕

小西友七著; 英語の前置詞 大修館, 1976  
D.J. ALLERTON; VALENCY AND THE ENGLISH VERB ACADEMIC PRESS  
H.T. WOOD; ENGLISH PREPOSITIONAL IDIOMS MACMILLAN PRESS

『表4 - between と先行詞の意味関係リスト (部分)』

(1) 「関係」

NP between A and B/ NP between C

調和	agreement, compatibility, coincidence, interaction, interworking, correlation, relation, relationship, interconnection, collaboration, cooperation, pair, connection, link, equivalence, alternation, interchange, transition, transmission
	回路的 circuit, path, point, interface, channel, cable, distance
不調和	interruption, decoupling, division, loss, imbalance, difference, crosstalk, distortion

(2) 「関係」

V between A and B/ V between C

differentiate, distinguish, select

\*\* 前置詞句の述部表現の中心省略句 (Hamby Code) VP 3a のこと

(3) 「関係」

VP between A and B/ VP between C

agree, connect, divide

\*\* 前置詞句の省略句

\* \* the relative phase between the modulated carrier and the pilot carrier

liaison point between countries のように NP 内の 2 名詞を結ぶ、修飾語句のため「関係」の意味のある例もいくつか見られる。理花、扇尾が示すように、NP と VP の間に - a もの区別が対称的で、この区別、この区別をなくす対象で、この区別。

\* \* between の意味関係リストは、「関係」以外に、「時間」、「場所」があり、それぞれが先行詞を結ぶ、  
「時間」 - time, period, interval, duration, pause, delay

「場所」 - region, range, band; vary, range; insert, introduce, interpose

区別がある。