

日本語技術文における並列構造†

首藤 公昭** 吉村 賢治** 津田 健蔵**

自然言語を工学の対象とする研究においては、言語現象を工学の立場から精密に調査・分析し、処理に有効な情報を整理する段階が基本的に重要である。このような認識に基づき、本論文では、自然言語処理で一つの障壁となっている並列構造に関して、約1万の日本語技術文を対象として行った調査の結果を報告する。約3,500か所に現れた並列構造を網羅的に抽出し、文形式の並列、名詞句の並列、その他の並列の3種に分類したうえで、それぞれどのような型があるかを出現頻度もあわせて明らかにする。これによって、日本語技術文における並列化現象の全体に見通しを与える。また、各型の並列構造に特有の共起表現とその出現頻度、および束ねの用法における共起表現の種類を明らかにし、機械処理における並列構造認定のための手がかりを与える。並列要素間の類似性を検出する手がかりとして、並列要素が同一文字を含む場合についての調査結果の一部も示す。以上の調査は、主に表層上の現象に関するものであり、言語の意味理解をただちに可能とする十分な情報を与えているわけではないが、言語を機械処理向きに標準化する問題や、将来の広範な日本語処理のための基礎的な資料として有効であると考えられる。

1. はじめに

自然言語の機械処理の研究においては、対象である言語に関する体系的研究が相当に不足しているため、言語現象および意味の世界を現在扱っている狭い領域から、より広範で“自然”な領域へと徐々に拡大し、システムを改良していくアプローチが基本的に要求される。しかしながら、自然言語の膨大さから、この方法には、度重なる改変を行っていくうちにシステム全体の明晰さが失われ、開発の続行が困難となる危険性がある。したがって、本格的な自然言語処理システムを開発するためには、あらかじめ、自然言語の本質的な特徴である表現形式や意味内容の多様性にできる限りの配慮を行っておくことが必要である。このためには、大量データに基づいて機械処理の立場から言語現象を調査し、有効な情報を体系的に分類・整理しておくことが大切であるが、従来、この種の基礎研究は十分には行われていない。本論文では、以上の認識に基づいて行った大量日本語データを対象とした並列構造に関する調査結果を報告する。広範な言語現象に対処しうる日本語機械処理を実現するための基礎的な資料を提供することが本論文の目的である。

一文中に同等の機能をもつ複数個の文要素を並べた構造を並列構造（または、等位構造）と呼ぶが、この種の構造は文要素の省略とも深く絡んで言語構造に異質なものを持ち込むため、従来、言語学においても、

難しい問題とされている⁸⁾⁻¹⁵⁾。

日本語文における並列構造に関する研究としては、助詞の連言的意味を英語と比較して論じた国広の研究¹⁾、並列要素の意味上の類似性について考察した石綿の研究²⁾、語の並列を調査した水谷、田中の研究³⁾、および小野原の研究⁴⁾、機械処理において並列構造を認定するための手がかりを並列名詞句と連用中止法について調べた長尾らの研究⁵⁾、などがあるが、いずれの場合も考察や調査の範囲が限定されているのが実情である。

本論文で示す調査結果は、技術雑誌（「日経コンピュータ」誌等）に載った42編の解説文（10,211文、約22万語）に基づくもので、主として並列構造の表層的な型を対象とするが、名詞句の並列から、文の並列に至る並列現象の全領域に目を向けた網羅的な調査となっている点に大きな特徴がある。

以下、2章では、並列の単位となる語列によって並列構造がどのように分類されるかを示す。3章および4章では、2章の分類に基づき、並列の型、並列構造に共起しやすい表現の調査結果を示す。3章は文形式の並列に関する結果、4章は名詞句の並列および非名詞句の並列に関する結果である。また4章では名詞句の並列のうち、特に束ねの用法に関する調査結果も示す。

2. 並列要素

並列の単位となっている語列を並列要素と呼ぶ。並列要素の種類には種々のものがあり、細かく列挙することは難しいが、大きくは、以下のように分類される。

† Coordinate Structures in Japanese Technical Sentences by KOSHO SHUDO, KENJI YOSHIMURA and KENZO TSUDA (Department of Electronics, Faculty of Engineering, Fukuoka University).

** 福岡大学工学部電子工学科

- { I 文形式
 II 未完形式 { i 名詞句
 ii 非名詞句

文形式とは末尾に述部を備え、まとまった叙述機能をもつ文節列^{6),7)}とする。日本語の特徴として文形式の文末側から取った文節列で、述部を含んだものは再び文形式である。また、文形式から述部を含む文末側部分語列を取り去って残った語列のうち、文形式でないものを未完形式と呼ぶ。未完形式は単独の名詞句の場合とその他の場合とに分かれる。

たとえば、例文(1)は文形式であり、その部分列が上記のいずれに分類されるかを(2)~(8)に例示する。これらの部分列がいずれも並列要素となりうる点に注意する。

- (1) 内部ノードの電位信号波形を 1 ns の

- 分解能で測定する I
 (2) 測定する. I
 (3) 1 ns の分解能で測定する I
 (4) 1 ns の分解能 II-i
 (5) 内部ノードの電位信号波形 II-i
 (6) 電位信号波形を 1 ns II-ii
 (7) 1 ns の分解能で II-ii
 (8) 内部ノードの電位信号波形を 1 ns の II-ii
 以下、上記の分類に従って調査結果を示す。

3. 文形式の並列

3.1 並列の型

文形式の並列のおもな型とその出現頻度を表1に示す。表中、文形式をSと記す。文形式と文形式の接続は意味上、多岐にわたっており、大量の資料の個々の

表1 文形式の並列の型
Table 1 Types of coordination of sentential forms.

項 数	型	出 現 数	出現比率	後置表現との共起
2 項	S [用], S	(300)	49.8%	(28)
	S [止] し (,) S	31	5.1	0
	R [用], かつ (,) S	23	3.8	0
	S [用] たり (だり) (,) S	20	3.3	2
	S [用] たり (だり) (,) S たり (だり)	19	3.2	0
	S [用], また (,) S	19	3.2	0
	S [体] うえ (うえに) (,) S	17	2.8	0
	S [止] か (,) S か	15	2.5	0
	S [用], しかも (,) S	15	2.5	0
	S [体] のに対し (,) S	14	2.3	0
	S [用], さらに (,) S	10	1.7	0
	S [止] と共に (,) S	9	1.5	0
	S [体] だけでなく (,) S	7	1.2	0
	S [止] と同時に (,) S	6	1.0	0
	S [止] か (,) あるいは (,) S	6	1.0	0
	S [体] いっぽう (,) S	6	1.0	0
	S [止], S	5	0.8	5
	S [体] ばかりでなく (,) S	3	0.5	0
	S [体] ばかりか (,) S	3	0.5	0
	S [用] (,) そして (,) S	1	0.2	0
	その他	25	4.2	1
3 項	S [用], S [用], S	(20)	3.3	(0)
	S [止] か (,) S [止] か (,) S か	6	1.0	1
	その他	7	1.2	3
4 項以上	S [用], S [用], S [用], S	8	1.3	0
	その他	7	1.2	1
総 計		(602)		(41)

[用], [止], [体] は、それぞれ、連用、終止、連体形を示す。
 (,) は読点が存在する場合と存在しない場合があることを示す。
 括弧内の数値は概数を表す。

事例について“等位”の接続であるか否かを正しく判定するのは難しいが、ここでは、並列要素の順序を入れかえ、必要に応じて表現の小変更（「も」を「が」に換えるなど）を行った場合に全体の意味に大きな差が生じないことを規準として選択した。

連用中止による接続は計 995 か所に認められ、そのうち 300 か所程度が等位と判定された。

いわゆる体言止めで後続の文形式と並置した形式は、接辞「だ」、「である」等に叙述機能があるものと考えて、未完形式の並列とみなす。並列要素間に介在して並列関係を指示する表現は 34 種確認された。

(9)～(12)に文形式の並列構造の例を示す。以後、並列要素を下線、並列関係を下線をつなぐ弧によって表す。

(9) 高速で、高集積な LSI が……

(10) 硅化物を P 領域表面に形成したり、P 層を形成せず、金属あるいはシリサイドで直接ショットキ接合を作り、ソース、ドレインとするなどの方策が…

(11) P チャネル MOS のソースはシリサイドを使い、同時にゲートを多結晶 Si とシリサイドの層とする方針をとった。

(12) 使用温度範囲は米軍規格なみに広いし、電源電圧変動も大きい。

3.2 後置される表現

例文(10)における「などの」のように、並列構造の直後に置かれて、並列構造全体を体言化したり、他の類似の事象を暗示する表現がある。これらは機械処理において並列のスコープを判定する際の手がかりを与える。調査の結果、この種の表現は「など」、「などの」、「などといった」、「などという」、「と」、「といった」、「という」の 7 種が認められた。これらの表現の使用頻度も表 1 に併記する。

4. 未完形式の並列

4.1 名詞句の並列

4.1.1 並列の型

名詞句の並列のおもな型とその出現頻度を表 2 に示す。表中、名詞句を NP と記す。並立助詞や並立の接続詞を含め、名詞句の並列を指示する表現、35 種が認められた。

2,675 か所の並列名詞句のうち、約 81% が 2 項の並列であり、その約 42% が <NP と NP> 型である。<(NP,) ⁱ および NP, $i=2\sim 4$ > 型が 31 例あるの

に対して、表 2 中には示していないが、<NP および NP (, NP) ⁱ> 型は $i=2$ のときの 2 例だけであり、「および」を用いる際は英語的表現が優勢であった。頻度は低いが「あるいは」、「または」、「それに」、「さらに」の場合も同じ傾向が見られた。

名詞句の並列構造の例を(13)～(18)に示す。

(13) パシベーション膜として、ポリイミド系樹脂や Al₂O₃ などがある。

(14) 一つの登録パターンと入力パターン間の距離が…

(15) n チャネルおよび P チャネル MOSFET 自身を微細にする場合の問題とウェルを微細にする場合の問題がある。

(16) データ処理と信号処理能力を…

(17) VTR のサーボ制御 LSI やデジタル・オーディオの信号処理用 LSI などがある。

(18) 航空機や軍用電子機器に…

機械処理においては、並列要素間の類似性を検出し、並列構造を確定する処理が必要となる。類似性には種々のレベルが考えられ、どのような枠組みで対処するかが今後の課題であるが、最も表層的な類似性として、例(14)～(18)のように、並列要素に共通の語や文字が現れる場合がある。この種の類似性の検出は比較的容易であり、ある程度の効果が期待できる。2 項の並列の主な型について、この種の類似性がどのくらい認められたかを表 3 に示す。これらの型では、平均して 1/3 の場合に同一文字の出現が認められた。

4.1.2 束ねの表現

例文(15)は、たとえば、「n チャネル…する場合の問題がある。そして、ウェルを…する場合の問題がある。」と 2 文に分解して言い換えることができる。これに対して、例文(14)では、「一つの登録パターン間の距離が…。そして、入力パターン間の距離が…」という言い換えは原文の意味を損うため不可能である。このように、名詞句の並列には、名詞句を単に列挙し、その他の共通な文要素を重複して表現するのを避ける用法と、名詞句を並置することによって新しい(複合的)概念を表示する用法とがある。本稿では後者を水谷ら³⁾にならって“束ね”と呼ぶことにする。一般に並列を含む文が与えられたとき、どちらの用法で並列構造が使われているのかを正しく判定するためには語用論的処理が必要となる場合もあり、今後の研究課題であるが、例文(14)の接尾語的表現「間」や動

表 2 名詞句の並列の型
Table 2 Types of coordination of noun phrases.

項 数	並 列 の 型	出 現 数	出 現 比 率	束 ね	後置表現との共起
2 項	NP と (,) NP	904	33.8%	288	0
	NP や (,) NP	545	20.4	0	108
	NP, NP	297	11.1	18	62
	NP (,) および (,) NP	142	5.3	7	3
	NP (,) あるいは (,) NP	85	3.2	0	6
	NP と (,) NP と	65	2.4	47	0
	NP のほか (に) (,) NP	15	0.6	0	0
	NP に加えて (,) NP	13	0.5	0	0
	NP (,) もしくは (,) NP	12	0.4	0	0
	NP・NP	11	0.4	0	1
	NP と共に (,) NP	10	0.4	0	0
	NP だけでなく (,) NP	10	0.4	0	0
	NP (,) または (,) NP	6	0.2	0	0
	NP (,) さらに (,) NP	6	0.2	0	1
	NP (,) ないし (,) NP	5	0.2	0	0
	NP か (,) NP か	5	0.2	0	0
	NP (,) かつ (,) NP	5	0.2	0	0
	NP か (,) NP	3	0.1	0	1
	NP とか (,) NP	3	0.1	0	1
	NP (,) そして (,) NP	3	0.1	0	0
	そ の 他	27	1.0	0	7
3 項以上	NP (, NP) ⁱ $i=2\sim6$	202	7.6	18	79
	NP や (,) NP (, NP) ⁱ $i=1\sim5$	143	5.3	2	78
	NP と (,) NP (, NP) ⁱ $i=1\sim6$	75	2.8	17	12
	(NP,) ⁱ および (,) NP $i=2\sim4$	31	1.2	5	3
	NP と (NP,) ⁱ および (,) NP $i=1\sim3$	11	0.4	0	0
	(NP,) ⁱ そして (,) NP $i=2\sim3$	5	0.2	0	2
	(NP,) ⁱ あるいは (,) NP $i=3, 4$	3	0.1	0	2
	そ の 他	33	1.2	3	1
総 計		2, 675		405	367

表 3 同一文字の出現比率
Table 3 Occurrence rate of identical characters
in coordinated expressions.

並 列 の 型	総出現数	並列要素に共通 文字のあるもの
NP と NP	904	314 (34.7%)
NP や NP	545	125 (22.9%)
NP, NP	297	153 (51.5%)
NP (,) および NP	142	58 (40.8%)
NP (,) あるいは NP	85	25 (29.4%)

詞「組み合せる」、形容詞「等しい」などと共起した並列構造は、束ねの用法と解釈しうる場合が多い。し

たがって、束ねと共起するこれらの表現を整理しておくことは、機械処理において重要である。調査の結果、表 4 に示すような表現が認められた。また、表 2 には、並列の各型について束ねの用法の頻度も示している。2 項の並列<NP と NP>型の約 32%、<NP と NP と>型の約 72%が束ねの用法であり、この差は、小野原の調査結果⁴⁾と同じ傾向を示している。

表 4 では、上記の表現の並列名詞句との係わりに基づいて分類した結果を示している。これらの文法カテゴリーの間には、図 1 に示すような形態論的遷移関係が見られる。表 4 の分類情報と図 1 のルールは、構文

表 4 「束ね」を求める表現とその分類

Table 4 Expressions which require *tabane*-type (bunch-type) coordination and their grammatical categories.

並列構造との係わり方	文法的分類	表 現 例
「が」を介して並列名詞句と結合するもの	動 詞 G-V	つり合う, つながる, 重なる, 重なり合う, 分れる, 似る, 違う, 会う, ...
	形容詞 G-A	等しい, ...
「を」を介して並列名詞句と結合するもの	動 詞 W-V	かみ合せる, 組み合せる, みわけ, 結ぶ, つなぐ, 切り離す, 足す, 併せ持つ, つき合せる, 対応づける, ...
「に」または「へ」を介して並列名詞句と結合するもの	自動詞 N-IV	分れる, ...
	他動詞 N-TV	分ける, 切り離す, ...
「から」または「より」を介して並列名詞句と結合するもの	動 詞 Y-V	成る, 成り立つ, ...
「の」を介して並列名詞句と結合するもの	サ変名詞 G-NV	対立, 一致, 交信, 重畳, 接触, 分離, 共有, 協力, 混在, 両立, 協調, 提携, ...
	サ変名詞 W-NV	集積, 結合, 結線, 比較, 識別, 併用, 弁別, 分離, 合成, 絶縁, 共載, 接合, 並列処理, ...
	名 詞 NO-N	あいだ, 距離, 中間, 間隔, 境界, インタフェース, 差, 和, 兼ね合い, ...
「への」を介して並列名詞句と結合するもの	サ変名詞 N-NIV	分化, 分極, ...
	サ変名詞 N-NTV	分離, 大別, 分類, クラス分け, ...
「よりの」または「からの」を介して並列名詞句と結合するもの	サ変名詞 Y-NV	成立, ...
並列名詞句に「の」を介して, または直接結合するもの	名 詞 φ-N	両方, 一方, 片方, いずれか, いずれの, いずれも, どちらか, どちらの, どちらも, 各~, 両~, [数量表現], ...
並列名詞句に直接結合するもの	接尾語 SFX	共, 間, ...
その他	形容動詞 G-AJV	緊密, 同一, 同じ, 独立, 一体, 同時, 別々, 対称, 類似, 共通, 同等, 交互, 同~, 等~, 並列, 異~, ...
	副 詞 ADV	それぞれ, 各々, 交互に, 相互に, 共に, ...

解析のための情報として有効である。

4.1.3 後置される表現

4.1.2 項で述べた共起表現のほか, たとえば, 例文(13)や(17)における「など」のように, 並列句の直後に置かれ, 並列スコープの判定に有用な表現がある。これらは, 「など」, 「などの」, 「などといった」, 「といった」, 「などのような」, 「などのように」, 「のよう」, 「のよう」の8種であった。この種の表現の出現頻度を表2に併記する。〈NP や NP〉型の場

合, 約 19.8%, 〈NP, NP〉型で約 20.9%, 〈NP (, NP)〉型で約 39.1%, 〈NP や NP (, NP)〉型で約 54.5% にこの種の表現が共起している。

4.2 非名詞句の並列

ここで対象とする並列要素は, 文形式から右端の部分語列を述語込みで取り去った残りの語列で, 単一の名詞句と見なせないものである。この種の要素からなる並列については, 従来, 研究の蓄積が少ない。変形文法では, 例文(19)のような英語における空所化

表 5 非名詞句の並列の型
Table 5 Types of coordination of non-sentential forms other than noun phrases.

項 数	句 数	並 列 の 型	出現数
2 項	1 個	AJV, AJV	24
		AJV と (,) AJV	9
		AJV・AJV	6
		AJV (,) および (,) AJV	4
		AJV (,) あるいは (,) AJV	4
		AJV (,) かつ (,) AJV	3
		NP' にも, NP' にも	3
	その他		21
	2 個	NP は NP', NP は NP'	33
		NP NP', NP NP'	24
		NP が NP', NP が NP'	16
		NP を NP', NP を NP'	7
		NP に NP', NP に NP'	7
		NP は NP' で, NP は NP' で	5
		NP で NP', NP で NP'	4
3 項	1 個	NP を NP' に, NP を NP' に	4
		その他	37
	2 個	2 項の関係表現パターンが同じもの	4
		不揃いのもの	4
		NP が NP で NP', NP が NP で NP'	1
		NP が NP で NP', NP が NP で NP'	1
		NP を NP', NP を NP', NP を NP'	1
		その他	3
	3 個	NP が NP で NP', NP が NP で NP'	1
		その他	4
	総 計		243

目し、日本語技術文、約 1 万を対象とした調査結果の一部を報告した。特に、語の並列、文の並列だけでなく、中間レベルの並列についても網羅的な調査を行い、日本語技術文に現れる並列構造の全体像を明らかにしようとした。表 1, 2, 5 によって、10,211 文中に現れた 3,520 か所程度の並列構造の型が覆われているが、その他の並列と見られる構造は、例文 (29)~(34) 等の非対称性をもつもの十数例、および (35)~(37) に示す例外的なもの数例にすぎず、日本語技術文の並列構造の 99% 以上が類型化できた。

(35) A 点と B 点の電圧を、それぞれ、-5V, -3V とする。

(36) 電卓しかり、時計しかり…

(37) 解説であれ、寄稿論文であれ、…

本稿の調査結果は並列構造自体の表層的な型に主眼をおいたものであるが、並列構造が文中でいかに使われ、並列を含む一般的な日本語の構造がどのようなものかを包括的に明らかにする必要がある。この点についても大量データに基づく調査・研究を進めており、稿を改めて報告したい。

並列構造を機械処理の対象とするためには、要素間の類似性を取り扱うことが不可欠である。本稿では、主として、表層上の手がかりを与えたが、さらに構文・意味的観点から類似性の研究を行う必要があり、今後の課題である。また、並列が“束ね”の用法であるか否かという意味解釈上の基本的な問題がある。これ

については、4.1.2 項で若干の手がかりを与えたが、今後、語用論まで含めた研究の蓄積が必要である。

本研究は、一部、放送文化基金(昭和56年度後期)、文部省科学研究費(特定研究「言語の標準化」)、および福岡大学言語工学研究専門委員会研究費によった。

謝辞 本研究は、資料の収集を担当していただいた、中野奈保子、宮永恵子、平田和子、加治明美、成清仁美、村上英子、野上富世の諸氏のご協力に負うところが大きい。ここに記して謝意を表します。また、有益な助言をいただいた福岡大学人文学部、久保善宏講師に感謝いたします。

参 考 文 献

- 1) 国広哲弥: (And) と「と・に・や・も」, 言語研究, No. 50, pp. 34-49 (1967).
- 2) 石綿敏雄: 構文解析自動化の研究 I, 国立国語研報, No. 34, pp. 139-174 (1969).
- 3) 水谷, 田中: 語の並列結合子, 計量国語学, No. 63, pp. 19-36 (1972).
- 4) 小野原生子: 語の接続, 東京女子大日本文学, No. 58, pp. 79-92 (1982).
- 5) 長尾, 辻井, 田中, 石川: 科学技術論文における並列句とその解析, 情報処理学会研資, 36-4, pp. 1-9 (1983).
- 6) 首藤, 檜原, 吉田: 日本語の機械処理のための文節構造モデル, 信学論, Vol. J 62-D, No. 12, pp. 872-879 (1979).
- 7) Shudo, K., Narahara, T. and Yoshida, S.: Morphological Aspect of Japanese Language Processing, COLING 80 Proceedings, pp. 1-8 (1980).
- 8) Chomsky, N.: *Syntactic Structures*, The Hague, Mouton (1957).
- 9) Gleitman, L.: Coordinating Conjunctions in English, in Reibel, D. and Schane, S. (ed.): *Modern Studies in English*, pp. 80-112, Prentice-Hall, Englewood Cliffs (1969).
- 10) Schachter, P.: Constraints on Coordination, *Language*, Vol. 53, No. 1, pp. 86-103 (1977).
- 11) Schachter, P.: A Note on Syntactic Categories and Coordination in GPSG, *Natural Language and Linguistic Theory*, Vol. 2, No. 3, pp. 269-281 (1984).
- 12) Lakoff, G. and Peters, S.: Phrasal Conjunction and Symmetric Predicates, in Reibel, D. and Schane, S. (ed.): *Modern Studies in English*, pp. 113-142, Prentice-Hall, Englewood Cliffs (1969).
- 13) Sjoblom, T.: Coordination, MIT Dissertation (1975).
- 14) Gazdar, G.: Unbounded Dependency and Coordinate Structure, *Linguistic Inquiry*, Vol. 12, No. 2, pp. 155-184 (1981).
- 15) Gazdar, G., Pullum, G.K., Sag, I.A. and Wasow, T.: Remarks and Replies, Coordination and Transformational Grammar, *Linguistic Inquiry*, Vol. 13, No. 4, pp. 663-676 (1982).

(昭和60年4月30日受付)

(昭和60年9月19日採録)