

# フランス語動詞補語の下位分類と 組合せ範疇文法による漸進的解析

高橋直人 竹内泉 一杉裕志

産業技術総合研究所人工知能研究センター

{naoto.takahashi,takeuti.i,y-ichisugi}@aist.go.jp

## 概要

本稿では、組合せ範疇文法を用いたフランス語二重補語構文の解析に関して考察する。文法的に正しい文のみを受理し、非文法的な文を排除するためには、補語にどのような下位分類が必要か分析し、更にそのような二重補語構文を人間同様に文頭から1語ずつ漸進的・逐次的に解析するための統語範疇および意味表示を提案する。

## 1 はじめに

人間同士はほとんどの場合に自然言語を用いて意志疎通を行っている。したがって近い将来、人間と人工知能との間で意志疎通が行われる際にも、自然言語がそのベースになると考えられる。特に、人間が人工知能に対して非定型的な質問や命令を発する際には、もっぱら自然言語が用いられる可能性が高い<sup>1)</sup>。

そのような意志疎通を可能にするためには、自然言語が時として示す一見不合理なまでに複雑な文法規則を、人工知能が十分に理解し、人間と同じように解釈する必要がある。更に、人工知能がそのように理解し解釈していると、人間が思えることが重要である<sup>2)</sup>。そのためには、文法と意味との対応が明確に定義されていなければならない。

また、人間が自然言語を理解する際は、文の終了を待たず、文頭から漸進的に解析を行っていると考えられ、その推測を支持する心理学的実験結果も存在する [1, 2]。更に工学的な応用として同時通訳や字幕生成などの自動化を実現するためには、漸進的な解析技術は不可欠であると言える。

1) 他方、人工知能から人間への回答には、画像や音響信号などの非自然言語的な表現が用いられる状況も十分考えられる。

2) 大規模言語モデルの登場によって自然言語処理技術は格段に向上したが、大規模言語モデルが何をどのように理解し解釈しているかは、本稿執筆時点ではまだ解明されていない。

本稿では複雑な文法規則とその意味の対応例として、フランス語二重補語構文における語順を取り上げる。次に、フランス語二重補語構文を組合せ範疇文法を用いて解析する際に必要となる補語の下位分類を設定し、更にそのような二重補語構文を人間同様に文頭から1語ずつ漸進的・逐次的に解析するための統語範疇および意味表示を提案する。

## 2 フランス語の動詞補語

### 2.1 語順と制約

フランス語は SVO 型の言語とされているが、動詞の補語（目的語）は代名詞化されると動詞の前に移動する。このとき、直接補語と間接補語の両方が代名詞化された場合の語順や、同時使用可能な代名詞の組合せには各種の制約がある [3]。

たとえば“Jean présente ses collègues à Marie.”（ジャンは彼の同僚たちをマリーに紹介する）という文における直接補語 ses collègues と間接補語 Marie を共に代名詞化した場合は、図 1 上部に示すように直接補語 les が間接補語 lui の前に来るが、もし間接補語が me（私に）であった場合は、反対に間接補語が直接補語の前に来なければならない（図 1 下部）。

また、直接補語としての me, te, nous, vous と、

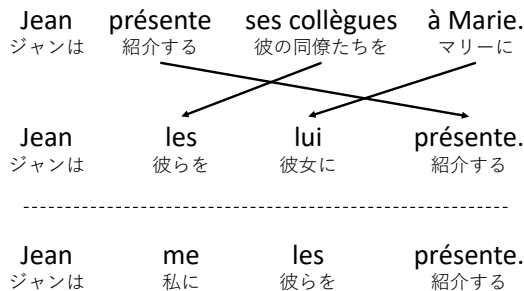


図 1 代名詞化された補語の語順。図上部では代名詞化された間接補語 lui が直接補語代名詞 les の後に置かれるが、図下部では間接補語代名詞 me が直接補語代名詞 les の前に置かれる。

間接補語の *lui*, *leur* を同時に用いることはできないという制約がある (表 1)。

したがってフランス語解析のための文法を設計する際には、通常の英語解析の際とは異なり、代名詞の統語範疇を一律に名詞句とすることはできない。もしそうすれば、非文法的な文をも受理してしまうという問題が生じるからである。

## 2.2 補語の下位分類

文法的に正しいフランス語のみが受理されるようにするため、本稿ではフランス語の動詞補語を、その文中の出現位置に基づいて、表 1 のように 4 種類に下位分類する<sup>3)</sup>。これら 4 種類の補語と動詞 *V* が文の表層上で出現する順序は

$$NP_m \rightarrow NP_l \rightarrow NP_u \rightarrow V \rightarrow NP_n \quad (1)$$

となる。

またこれら 4 種類の動詞補語は、表 2 に示すように、直接補語代名詞として使用可能か否か (*D*)、間接補語代名詞として使用可能か否か (*I*)、という 2 種類の素性に分解して解釈することができる。上記 (1) で示した出現順序を素性形式で表現すると

$$\left[ \begin{array}{c} +D \\ +I \end{array} \right] \rightarrow \left[ \begin{array}{c} +D \\ -I \end{array} \right] \rightarrow \left[ \begin{array}{c} -D \\ +I \end{array} \right] \rightarrow V \rightarrow \left[ \begin{array}{c} -D \\ -I \end{array} \right] \quad (2)$$

となる。

## 3 組合せ範疇文法

組合せ範疇文法 [4, 5, 6, 7] は語彙化文法の一環であり、古典的範疇文法に関数合成規則を取り入れたものである。「マイルドな文脈依存」と言われる弱生成能力を持つため、文脈自由文法の範囲に収まらない自然言語現象を記述する事ができる。また理論的にも実際的にも効率のよい解析アルゴリズムが知られており [8, 9]、自然言語の文法記述に適しているとされる。

組合せ範疇文法では、*S* (文) や *NP* (名詞句) のような少数の基底範疇と、それらを結ぶ演算子を再帰的に用いる事で、多様な統語範疇を表現する。演算子には / と \ がある。*X* と *Y* を統語範疇としたとき、*X/Y* は「右側に *Y* が隣接したときに *X* となる統語範疇」、*X\Y* は「左側に *Y* が隣接したときに *X* となる統語範疇」を意味する。

3) 動詞補語としては表 1 以外にも、前置詞を伴って動詞の後に置かれる人称代名詞強勢形や、動詞の前に置かれる中性代名詞等があるが、それらの分析に関しては別の機会に譲る。

表 1 文中の出現位置に基づくフランス語補語の下位分類。なお、一つの動詞が *NP\_m* と *NP\_u* を同時に取ることはできない。

記号	説明
<i>NP_m</i>	1・2 人称の直接・間接補語代名詞 (動詞の前, <i>NP_l</i> の前) 例: <i>me</i> (私を/私に), <i>te</i> (君を/君に), <i>nous</i> (私たちを/私たちに), <i>vous</i> (あなたたちを/あなたたちに)
<i>NP_l</i>	3 人称直接補語代名詞 (動詞の前, <i>NP_m</i> の後, <i>NP_u</i> の前) 例: <i>le</i> (彼を), <i>la</i> (彼女を), <i>les</i> (彼らを/彼女らを)
<i>NP_u</i>	3 人称間接補語代名詞 (動詞の前, <i>NP_l</i> の後) 例: <i>lui</i> (彼に/彼女に), <i>leur</i> (彼らに/彼女らに)
<i>NP_n</i>	代名詞以外の補語名詞句 (動詞の後) 例: 限定詞+普通名詞 ( <i>ses collègues</i> ), 固有名詞 ( <i>Jean, Marie</i> ), etc.

表 2 下位分類された補語の素性分解

		間接補語代名詞の使用 ( <i>I</i> )	
		+	-
直接補語 代名詞的 使用 ( <i>D</i> )	+	<i>NP_m</i> <i>me, te, nous, vous</i>	<i>NP_l</i> <i>le, la, les</i>
	-	<i>NP_u</i> <i>lui, leur</i>	<i>NP_n</i> 代名詞以外

各語彙項目に付与された統語範疇は、少数かつ言語非依存の組合せ規則によって互いに組み合わせられるが、このとき意味表示も同時並列的に合成される。この操作が再帰的に実行される事によって、語彙項目レベルの意味表示が文レベルの意味表示へと順次導出される。

本稿で使用する組合せ規則を図 2 に示す。コロンの左側が統語範疇で、右側がラムダ項による意味表示である。

### 3.1 二重補語動詞の語彙項目

組合せ範疇文法におけるフランス語二重補語動詞の語彙項目は、その直接補語と間接補語のそれぞれが表 1 に示したいずれの下位分類に属するかによって、表 3 のように 7 種類に分類される。

表 3 の文型 1 と文型 5 は、統語範疇の第 4 要素

表 3 二重補語動詞 *présente* の文型別語彙項目。文型欄における, S, V, COD, COI はそれぞれ主語, 動詞, 直接補語 (complément d'objet direct), 間接補語 (complément d'objet indirect) を意味する。

No	文型 統語範疇	例文 意味表示
0	S V COD COI $S \backslash NP_n / PP / NP_n$	Jean présente ses collègues à Marie. $\lambda x. \lambda y. \lambda z. \text{introduce}'xyz$
1	S COD V COI $S \backslash NP_n / PP \backslash NP_l$	Jean les présente à Marie. $\lambda x. \lambda y. \lambda z. \text{introduce}'xyz$
2	S COI V COD $S \backslash NP_n \backslash NP_n \backslash NP_u$	Jean lui présente ses collègues. $\lambda x. \lambda y. \lambda z. \text{introduce}'yxz$
3	S COD COI V $S \backslash NP_n \backslash NP_l \backslash NP_u$	Jean les lui présente. $\lambda x. \lambda y. \lambda z. \text{introduce}'yxz$
4	S COI COD V $S \backslash NP_n \backslash NP_m \backslash NP_l$	Jean me les présente. $\lambda x. \lambda y. \lambda z. \text{introduce}'xyz$
5	S COD V COI $S \backslash NP_n / PP \backslash NP_m$	Jean me présente à Marie. $\lambda x. \lambda y. \lambda z. \text{introduce}'xyz$
6	S COI V COD $S \backslash NP_n / NP_n \backslash NP_m$	Jean me présente ses collègues. $\lambda x. \lambda y. \lambda z. \text{introduce}'yxz$

$NP_l$  と  $NP_m$  が違うだけなので, 素性表現とワイルドカード \* を用いて [+D, \*I] と一つにまとめることができる。同様に文型 2 と文型 6 も, 統語範疇の第 4 要素  $NP_u$  と  $NP_m$  が違うだけなので, [\*D, +I] とまとめることができる。

### 3.2 その他の語彙項目

本稿の例文に登場する語彙のうち, 二重補語動詞 *présente* 以外の語彙項目を表 4 に示す。名詞句および前置詞句の意味表示は, それらの中に含まれる名詞の人称・数・性で表されるものとする。代名詞の意味表示に関しても同様に, その代名詞の人称・数・性で表されるものとする。

順関数適用規則	$\frac{X/Y:f \ Y:a}{X:fa} >$
順関数合成規則 ( $n = 1$ )	$\frac{X/Y:f \ Y/Z:g}{X/Z:\lambda x.f(gx)} > B$
順関数合成規則 ( $n = 2$ )	$\frac{X/Y:f \ (Y/W)/Z:g}{(X/W)/Z:\lambda x.\lambda y.f(gxy)} > B^2$
順型繰り上げ規則	$\frac{X:a}{T/(T \setminus X):\lambda f.fa} > T$

図 2 本稿で使用する組合せ規則一覧

表 4 二重補語動詞以外の語彙項目。Jean から collègues までの意味表示のうち, 先頭の数字は人称を, 2 文字めは数 ( $s$ : 単数,  $p$ : 複数) を, 3 文字めは性 ( $m$ : 男性,  $f$ : 女性,  $c$ : 男女共通) を表す。最後 2 語の *ses* と *à* の意味表示は, 恒等関数 (すなわち引数の意味表示そのもの) となる。

音韻表示	統語範疇	意味表示
Jean	$NP_n$	$3sm'$
Marie	$NP_n$	$3sf'$
les	$NP_l$	$3pc'$
lui	$NP_u$	$3sc'$
me	$NP_m$	$1sc'$
collègues	$N$	$3pc'$
ses	$NP_n / N$	$\lambda x. x$
à	$PP / NP_n$	$\lambda x. x$

## 4 漸進的導出

前述のように, 自然言語解析には漸進性が重要な意味を持つ。表 3 に挙げた全例文は, 組合せ範疇文法を用いて厳密な意味で漸進的に解析できる。各例文の漸進的導出過程を, 付録の図 3 から図 9 に示す。どの文においても, 新たな 1 語が入力される度に, その時点で作成中の部分木に即座に接続されていることがわかる<sup>4)</sup>。

意味表示の導出に関しては途中の過程を省略してあるが, 表 3 と表 4 に示した語彙項目, および図 2 に示した組合せ規則に従えば, すべて機械的に曖昧性なく決定される。

補語を代名詞化する際に性別に関する情報が欠落する場合が一部あるが, いずれの文型においても正しい意味表示が得られていることがわかる。

## 5 関連研究

組合せ範疇文法をフランス語に適用したという報告は, 意外にもあまり多くない。

Delamarre[10] は, 組合せ範疇文法が内包する過剰生成の問題について触れ, それを解決するために組合せ範疇文法の拡張であるマルチモーダル組合せ範疇文法 [11] とカテゴリータイプ論理を併用し, フランス語補語代名詞の問題に取り組んでいる。

Lêら [12] は, フランス語組合せ範疇文法のためのスーパータグ付与アルゴリズムを提案し, それを

4) ただし, すべての導出が曖昧性なく決定論的に進むわけではない。たとえば例文 5 および例文 6 において “Jean me présente” を読んだ時点では, この *me* が直接補語なのか間接補語なのかを判定することはまだできない。

用いて Abeillé らが作成したフランス語句構造文法ツリーバンク [13] を組合せ範疇文法ツリーバンクに変換した。変換に当たっては、オリジナルの句構造文法ツリーバンクに加え、それを依存構造に変換した Candito らの依存構造ツリーバンク [14] [15] から得た情報も利用している。論文を読む限り、補語代名詞の統語範疇は一律に名詞句 *NP* と指定されており、下位分類はされていないように見える。

## 6 おわりに

フランス語二重補語構文に出現する代名詞、名詞句、前置詞句などの各種補語を分類し、組合せ範疇文法を用いて文頭から1語ずつ漸進的・逐次的に解析するための統語範疇および意味表示を提案した。

本稿では触れなかったが、1・2人称直接補語と3人称間接補語を併用する場合や、肯定命令文で補語代名詞を用いる場合には、人称代名詞強勢形と呼ばれる型式を用いる必要がある。更に、*en*, *y*, *le* などの中性代名詞が人称代名詞と併用される場合や、本動詞が別の動詞の不定詞を従える場合には、補語の語形や語順に関してまた追加の文法規則が必要となる。これらを統一的に扱えるようにするのが、当面の目標である。

学習者向けの文法書であれば、代名詞の可能な組合せを網羅的に列挙するだけで十分であるかもしれないが、もし本稿 2.2 節に示したように、文法の背後に潜在する素性を仮定し、各種の文法規則がそれら素性の組合せによって説明できるのであれば、よりエレガントな文法理論となるであろう。

またフランス語と同系統のスペイン語やイタリア語でも、代名詞化された補語は一般に動詞の前に移動するが、その際の語順や制約には言語ごとの差異がある。各言語における語彙項目の違いを素性の違いに還元できるか興味を持たれる。

## 謝辞

本研究は JSPS 科研費 JP22K12188 の助成を受けたものです。

## 参考文献

- [1] Vera Demberg. Incremental derivations in CCG. In **Proceedings of the 11th International Workshop on Tree Adjoining Grammars and Related Formalisms (TAG+11)**, pp. 198–206, Paris, France, September 2012.
- [2] Miloš Stanojević and Mark Steedman. CCG parsing algorithm with incremental tree rotation. In **Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)**, pp. 228–239, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.
- [3] 目黒土門. 現代フランス広文典. 白水社, 2000.
- [4] Mark Steedman. **The Syntactic Process**. The MIT Press, 2000.
- [5] Mark Steedman. **Surface Structure and Interpretation**. The MIT Press, 1996.
- [6] 戸次大介. 日本語文法の形式理論. くろしお出版, 2010.
- [7] 峯島宏次. 論理と文法. 数学セミナー, Vol. 59, No. 3, pp. 30 – 35, 2020.
- [8] Krishnamurti Vijay-Shanker and David Weir. Polynomial time parsing of combinatory categorial grammars. In **28th Annual Meeting of the Association for Computational Linguistics**, pp. 1–8, 1990.
- [9] Stephen Clark and James R. Curran. Wide-coverage efficient statistical parsing with CCG and log-linear models. **Computational Linguistics**, Vol. 33, No. 4, pp. 493–552, 2007.
- [10] Simon Delamarre. Exploration de la Grammaire Catégorielle Combinatoire Multimodale. Master’s thesis, Université du Québec, Département Mathématiques et Informatique Appliquées, Trois-Rivières, Canada, September 2011.
- [11] Geert-Jan Kruijff and Jason Baldridge. Multi-modal combinatory categorial grammar. pp. 211–218, 01 2003.
- [12] Luyôn Ngọc Lê and Yannis Haralambous. CCG Supertagging Using Morphological and Dependency Syntax Information. In **International Conference on Computational Linguistics and Intelligent Text Processing**, Vol. Springer Lecture Notes in Computer Science of **Actes de la Conférence Internationale de Linguistique et de Traitement Automatique des Langues Naturelles**, pp. 608–621, La Rochelle, France, April 2019. Springer.
- [13] Anne Abeillé, Lionel Clément, and Alexandra Kinyon. Building a treebank for French. In M. Gavrilidou, G. Carayannis, S. Markantonatou, S. Piperidis, and G. Stainhauer, editors, **Proceedings of the Second International Conference on Language Resources and Evaluation (LREC’00)**, Athens, Greece, May 2000. European Language Resources Association (ELRA).
- [14] Marie Candito, Benoît Crabbé, and Pascal Denis. Statistical French dependency parsing: Treebank conversion and first results. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odijk, Stelios Piperidis, Mike Rosner, and Daniel Tapias, editors, **Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC’10)**, Valletta, Malta, May 2010. European Language Resources Association (ELRA).
- [15] Marie Candito, Benoît Crabbé, Pascal Denis, and François Guérin. Analyse syntaxique du français : des constituants aux dépendances. In Adeline Nazarenko and Thierry Poibeau, editors, **Actes de la 16ème conférence sur le Traitement Automatique des Langues Naturelles. Articles longs**, pp. 31–40, Senlis, France, June 2009. ATALA.



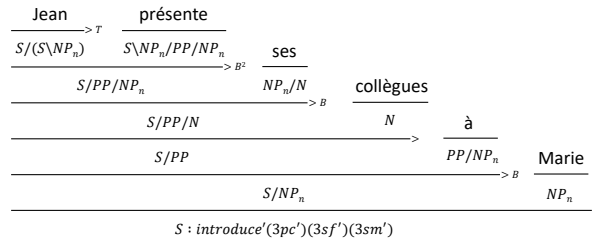


図3 例文0：直接補語と間接補語の両方が名詞の場合

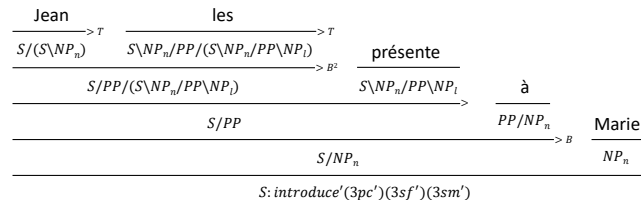


図4 例文1：直接補語が代名詞，間接補語が名詞の場合

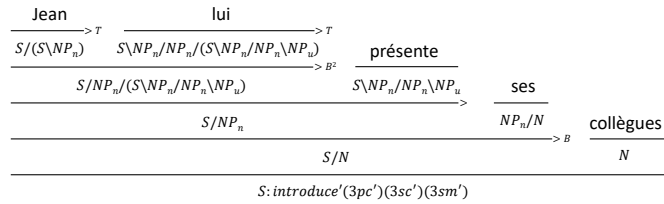


図5 例文2：間接補語が代名詞，直接補語が名詞の場合

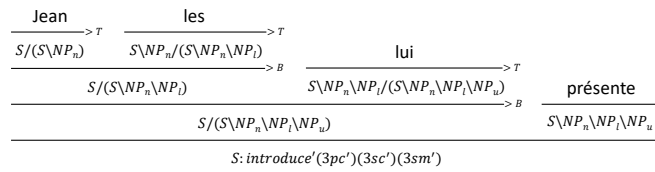


図6 例文3：直接補語代名詞が間接補語代名詞に先行する場合

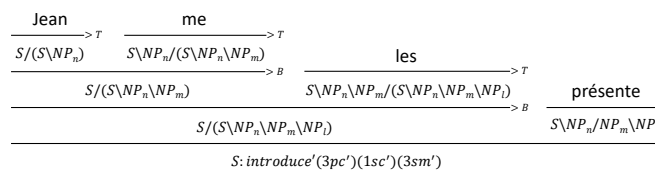


図7 例文4：直接補語代名詞が間接補語代名詞に後続する場合

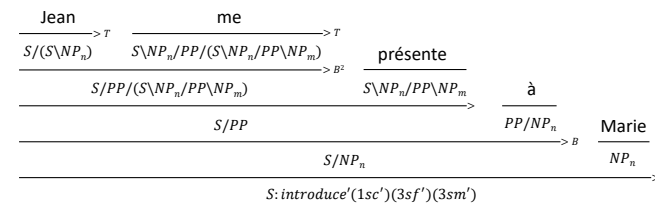


図8 例文5：直接補語が代名詞，間接補語が名詞の場合（その2）

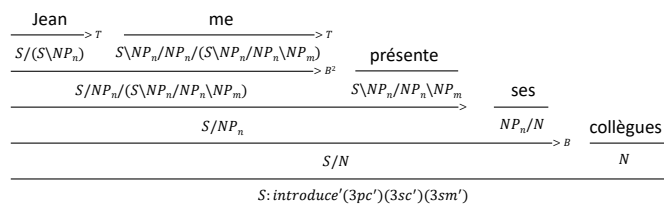


図9 例文6：間接補語が代名詞，直接補語が名詞の場合（その2）