

新聞コーパスの調査に基づくフランス語人称代名詞の使い分け基準について (“on”と“l'on”を例に)

清水正勝* 清水由美子** 赤間啓之*

概要: フランス語の人称代名詞である“on”と“l'on”は、意味的に等価なものとされている。17世紀の文法学者 Vaugelas は、両者の使い分けを「音」の観点から整理した。彼の解釈を後世の文法書も踏襲しているが、今回我々が行った新聞コーパスの調査によって、両者の使い分けは、彼の言う「音」という要素よりも、「単語そのもの」に依存する可能性が高いことを示す。

Linguistic and statistical research for newspaper corpora on the Web -Usage of the French impersonal subject pronoun, “on” (or “l'on”)-

Masakatsu Shimizu*, Yumiko Shimizu**, Hiroyuki Akama*

Abstract : The aim of this paper is to find out some practical criteria for choosing “on” or “l'on”, which are semantically equivalent as French impersonal subject pronoun. As for this alternative, a grammarian of 17th century, Vaugelas, proposed some constraints depending on the neighboring sounds or spellings and based on the easiness to pronounce, write, hear and read, which have come down to almost all the subsequent grammarians. But our statistical research for some newspaper corpora on the Web, using a chi-square test and the C5.0 algorithm, made us realize the effects of an unknown heterogeneous parameter for this alternative, “historical or psychological identity of word”, explicable from the viewpoint of ideological linguistics ascribed to Damourette and Pichon.

1. はじめに

1-1. 研究目的

フランス語の“on”は、漠然と「人々や我々」を指示する非人称主語のひとつであるが、この人称代名詞には、“l'on”という表記方法もある。この二つは、現代フランス文法においては、意味的に全く等価なものであるから、本来ならどの場面でもどちらを使用してもそれが許容されるはずであるが、実際には「標準形は“on”」という意識が働くためか、異形の“l'on”の出現が許される場面には、文法書によっていくつかの条件が与えられている。

赤間ら(2002)は、17世紀の文法学者 Vaugelas(1585-1650)が示した「制約」が、現代のフランス語での単語選択において、どの程度の守られているのかを、フランス新聞 Le Monde をコーパスとして調査し、人工知能エンジン C5.0 により分析したところ、彼の影響力は、現代でも十分に残っていることを示した。しかし、赤間ら(2002)のアプローチは、「母音衝突(hiatu s)」と「不快音調(cacophonie)」というきわめて巨視的な要素による調査であり、C5.0 というアルゴリズムが、単語選択における文法的な「制約」が、どのような場面で適用されるの

* 東京工業大学大学院 社会理工学研究科

** 武蔵工業大学 環境情報学部

* Graduate School of Decision Science and Technology, Tokyo Institute of Technology

** Faculty of Environmental and Information Studies, Musashi Institute of Technology

かを視覚化するために使用できることを示したに過ぎず、個々のケースを精査したわけではない。

また、20世紀の心理言語学者であり文法学者である Damourette と Pichon は、“on”と“l'on”を使い分けの際に適用される「制約」は、「音」という要素だけではなく、「単語そのもの」にも由来すると読みとれる研究を行っている。しかし Pichon らの主張は、赤間ら(2002)の「単語の最後が同一の音であるならば、それは同一の変数として扱う」という巨視的なアプローチでは、オミットされてしまうものであるため、彼らの主張を検証することもできない。

本論では、特に“l'on”に注目し、これが出現する全ての場面と文法書の記述との間の整合性を検証するとともに、Pichon らの主張と、Vaugelas の主張のどちらが優位であるのかを検証するために、「音」と「単語」に分け、場面毎に² 検定を行う。また、コーパス間で、“on”と“l'on”との使われ方に違いがあるのかどうかを調査するために、性格の異なる二つのフランス新聞をコーパスとして、“on”と“l'on”を中心語とする共起語の連なりである n-gram を全て抽出し、C5.0 を用いた分析に際しても、「音」で一括りにすることなく、「単語」として扱う変数も用意することで、より精緻で、現実を反映した決定木を描くことの可能性を検証する。

1-2. Vaugelas の提唱した「制約」

前節でも触れたように、“on”と“l'on”との間に意味的な相違は無いのであるが、実際の使用場面を調べてみると、完全に「使用者の自由」というわけではなく、“l'on”が出現するのはごく限られた範囲の中であることが分かる。これは、Vaugelas の影響が残っているためであると考えられる。彼は著書 *Remarques sur la langue française* の中で、「言いやすさ・書きやすさ・聞きやすさ・読みやすさ」に基づく「制約」を提唱し、そのほとんどが後の文法書へと受け継がれ今日に至っている。

彼が与えた、使い分けのための「制約」とは、「隣接する単語の文字(音)」との関係に依るものであり、要約すると以下の三点に絞られる。

直前の単語が母音字で終わっている場合には、「母音衝突」の発生を避けるために“l'on”を使用すること。

直後の単語が、“l”で始まっている場合には、「不快音調」を避けるために、“l'on”を使用しないこと。

同じ音が続かないように配慮して“on”と“l'on”を適宜使い分けること。

これらの「制約」が、「Vaugelas が示し、今日まで善くも悪くも続いてきた」(Dupré, 1972)のものであり、「“on”と“l'on”の使い分けを判断できるのは、使用者の『耳』だけである」(Littré, 1885)とされた所以である。

1-3. Damourette と Pichon の解釈

Pichon ら(1936)も“l'on”が使われる場面についての分析を行っているのだが、興味深いことに、彼らは「隣接する単語の音」との関係に注目するのではなく、「単語の歴史的な変遷」という独特の観点によって、それを論じているのである。「l'on」を好むものとして、以下の例を挙げている。

“si”の直後で“l'on”を使用するのは、文筆家や知識人である。

“l'on”は、昔“on”を“en”と表記していた地方で好まれる表記方法である。

前者は、“si”という単語が昔は「縮約」の対象であり、“on”が後続する場合には“s'on”と表記しなければならなかった事実を挙げ、「縮約」の対象でなくなっても、“si on”と表記することに抵抗がある文筆家や知識人が、“l'on”

を好んで使用しているという解釈である。

これは、Vaugelas らのような「母音衝突だから」とか「言いにくいから」とかいった観点からの解釈とは相容れないものである。

1-4. 本論のアプローチ法

本論では、前節で触れた Pichon らの解釈も踏まえ、「l'on」が出現する場面にはどのような傾向が見られるのかということに焦点を絞り調査することとした。方法としてはコーパスの中から、「on」あるいは「l'on」を中心語とする単語の連なりである n-gram を全て抽出し、出現比率の調査と C5.0 による分析を行うというものである。また、対象とするコーパスを複数とすることで、コーパス間での異同も調査した。

今回の調査によって、「on」と「l'on」の使い分けに関しては、Vaugelas の「音」という観点よりも、Pichon らが挙げたような「単語そのもの」という要素に依存している可能性が高いことが明らかになった。

2. 調査方法

2-1. 調査対象コーパス

今回の調査に使用したフランス新聞コーパスは下記の二紙である。

Le Monde (2000.01.01 ~ 2000.06.30) : 総単語数約 1,600 万語

Dernières Nouvelles d'Alsace (2000.01.01 ~ 2000.06.30) : 総単語数約 1,100 万語

Le Monde (以下 LeM と表記) は全国紙、Dernières Nouvelles d'Alsace (以下 DNA と表記) は地方紙であるという差があるが、それ以上に LeM のほうは、図や写真を掲載せずに、文章に関しても、「洗練された」言語表現にこだわる新聞とされ、知識人に向けて書かれている新聞である。

2-2. 5-gram の抽出と調査

新聞コーパスから、全ての「on」と「l'on」を中心語とする 5-gram を Perl5.0 のスクリプトにより抽出したものを使用する。1-1.でも触れているが、「l'on」が出現するのは、ごく限られた場面であるとされている。まず、「隣接する単語の文字(音)」という点に注目し、従来の文法書の記述がどれだけ守られているのかを調査する。次に、Vaugelas が「音」、Pichon らが「単語」という観点から、「l'on」の使用場面を規定している点に関して、どちらがより説得力のあるものかということをも、「音」と「単語」との双方に注目し、「l'on」の出現比率に統計的な差があるかどうかを調査・検定する。

2-3. C5.0 による決定木の生成

人工知能エンジン C5.0 は、SPSS 社が開発した Clementine という統計解析ソフトウェアに収録されているアルゴリズムのひとつであり、分析結果を決定木の形で表示することができる。これは、Quinlan が 1986 年に発表した ID3 (Iterative Dichotomiser 3) モデルで用いられている「利得基準」を改良した、「利得比基準」をもとにノードの設定と分岐を計算するものである。

これを使用することによって、「on」と「l'on」のどちらが使用される確率が高いのかということをも、隣接する単語(音)から推定することが可能になる。

3. 隣接する単語との関係の調査

3-1-1. “l'on”と直前の単語との関係

以下に，“l'on”の直前にどのような単語が頻出しているのかを調査した一覧を示す。(カッコ内は英語の逐語訳)

LeM (総パターン：4,951・異なりパターン：42)

DNA (総パターン：2,863・異なりパターン：58)

直前の単語	出現回数	占める割合 (%)
区切り文字 (, や .)	59	1.19
que (what)	2,593	52.37
si (if)	1,064	21.49
où (where)	701	14.16
et (and)	275	5.55
lorsque (when)	71	1.43
qui (who)	64	1.29

表 1 . LeM の“l'on”の直前に現れる主な単語

直前の単語	出現回数	占める割合 (%)
区切り文字 (, や .)	154	5.38
que (what)	1,473	51.45
où (where)	434	15.16
si (if)	385	13.45
et (and)	190	6.64
lorsque (when)	88	3.07

表 2 . DNA の“l'on”の直前に現れる主な単語

	区切り文字	母音字	子音字
出現回数	59	4,845	47
占める割合 (%)	1.19	97.86	0.95

表 3 . LeM の“l'on”の直前に現れる単語の最後の文字

	区切り文字	母音字	子音字
出現回数	154	2,663	46
占める割合 (%)	5.38	93.01	1.61

表 4 . DNA の“l'on”の直前に現れる単語の最後の文字

表 1・2 は, “l'on”が出現する場合の, 直前の単語 (出現比率 1.0%以上のもの) を列挙したものである。英語の “and”にあたる “et”の末尾の “t”は, 子音字として機能しない “t”であるため, 文法上「母音衝突」として扱われている。また, 表 3・4 では, 直前の単語の末尾が子音字か母音字かという観点から分布をまとめたものである。これらの表からわかるように, “l'on”が出現する条件として, 「母音衝突」が重要な要素であることがわかる。

3-1-2. “l'on”と直後の単語との関係

コーパス	直後が L	それ以外	合計
LeM	8(0.16)	4,943	4,951
DNA	4(0.14)	2,859	2,863

表 5. “l'on”が出現するときの次の単語の文字 (カッコ内は割合%)

“l'on”が出現する条件として, 直前の単語のみならず, 直後の単語も重要な要素となっている。Vaugelas 的な文法書では, 不快音調を回避することを目的として, 直後の単語が “l”で始まっている場合には, “l'on”を使用しないことを推奨している。表 5 が示すように, この「制約」が忠実に守られていることがわかる。

3-2-1. 発音が同一な場合 (“si”とそれ以外のイ音)

前節までの結果を見ると, Vaugelas の示した「制約」が忠実に守られており, 文法書の記述および赤間ら(2002)の研究が当を得たものであるという結論になる。しかし, 以下のクロス表と, 出現比率に統計的な差があるかどうかを検定する, ²検定の結果を見ていただきたい。

	on	l'on	合計
si	592	1,051	1,643
それ以外のイ音	88	83	171
合計	680	1,134	1,814

表 6. “si”と「それ以外のイ音」で見る “on / l'on” の出現回数 (LeM)

$$^2(1) = 15.74 > 6.63$$

	on	l'on	合計
si	713	383	1,096
それ以外のイ音	122	25	147
合計	835	408	1,243

表 7. “si”と「それ以外のイ音」で見る “on / l'on”の出現回数 (DNA)

$$^2(1) = 18.92 > 6.63$$

Vaugelas 的な判断に依れば, “si”と「それ以外のイ音で, かつ “i”で終わる単語」とは, 「母音で終わっている」という点でも, 末尾の文字という点でも差異がなく, クロス表でも似たような比率を示すことが予想される。しかし, ²検定の結果では 1%水準で有意差が出ており, 彼らの主張を支持することはできない。この “si”は Pichon らが「歴史的な経緯」から “l'on”が好まれると解釈している単語である。

3-2-2. 発音が同一な場合 (“ou”と“où”)

“ou”と“où”は、綴りにおいてアクセント記号の有無という差こそあれ、発音は全く同一の単語である。(カッコ内は英語の逐語訳)

	on	l'on	合計
où (where)	182	676	858
ou (or)	18	10	28
合計	200	686	886

表 8 . “où”と“ou”でみる“on / l'on”の出現回数 (LeM)

$$\chi^2(1) = 28.78 > 6.63$$

	on	l'on	合計
où (where)	200	419	619
ou (or)	17	3	20
合計	217	422	639

表 9 . “où”と“ou”でみる“on / l'on”の出現回数 (DNA)

$$\chi^2(1) = 23.98 > 6.63$$

ここでも、使用される比率に 1%水準で有意差が出ている。この二つの単語と“on”と“l'on”の関係に対する言及はどの文法書にも記述がないが、差が出たということは、今後、あらゆる文法書で記述していく必要があるだろう。以上見てきたように、“on”が“l'on”かという選択に関しては、「音」だけでなく「単語」という要素も加味した解釈が必要であることが明確になったといえる。

4. C5.0 の出力結果

前章の結果を踏まえ、C5.0 へ渡す変数は、隣接する音に注目した赤間ら(2002)の設定したものに改良を加え、同一の発音であっても出現比率に差があった「単語」を個々に変数として設定した。これにより、“on”と“l'on”のどちらがより好まれているのかという点に関して、より精緻な決定木が描け、単語選択の因果関係がよりわかりやすくなった。ここではその一部を示す。

(直前が si) : [最頻値 : l'on]	
(直後が con / com) : [頻度 : 75, 73.3%]	l'on
(直後が L) : [頻度 : 97, 97.9%]	on
(それ以外) : [頻度 : 1,471, 67.6%]	l'on
(直前がそれ以外のイ音) : [頻度 : 171, 51.5%]	on

図 1 . LeM における“si”と「それ以外のイ音」の決定木

(直前が si) : [最頻値 : <u>on</u>]	
(直後が con / com) : [頻度 : 53, 58.5%]	<u>l'on</u>
(直後が L) : [頻度 : 39, 100.0%]	on
(それ以外) : [頻度 : 1,004, 64.9%]	on
(直前がそれ以外のイ音) : [頻度 : 147, 83.0%]	on

図 2 . DNA における“si”と「それ以外のイ音」の決定木

図 1・2 は、直前の単語が“si”の場合を描いたものであるが、LeM と DNA では好まれる単語が違うということがわかる（図中の下線部）。“l'on”を使用する際には、前後の単語や、その音に十分な注意を払う必要があることは既に触れたとおりである。また、“l'on”には、「格式張った」印象があるという研究もあり、LeM は DNA と比較して、より高い言語意識と形式意識があるといえるのではないだろうか。このように、同一の条件であっても、最終的な単語選択においてコーパス間で差が出る例が発見されたことは、今後、社会言語学的な調査においてもこの分析方法が有効に働く可能性を示せたことになる。

(直前が ou) : [最頻値 : <u>l'on</u>]	
(直後が con / com) : [頻度 : 15, 86.7%]	l'on
(直後が L) : [頻度 : 44, 93.2%]	on
(それ以外) : [頻度 : 799, 82.6%]	l'on
(直前が ou) : [頻度 : 28, 64.3%]	<u>on</u>

図 3 . LeM における“ou”と“ou”に関する決定木

(直前が ou) : [最頻値 : <u>l'on</u>]	
(直後が con / com) : [頻度 : 19, 78.9%]	l'on
(直後が L) : [頻度 : 28, 92.9%]	on
(それ以外) : [頻度 : 572, 70.3%]	l'on
(直前が ou) : [頻度 : 20, 85.0%]	<u>on</u>

図 4 . DNA における“ou”と“ou”に関する決定木

図 3・4 は、直前が“ou”あるいは“ou”であるときの決定木である。表 5・6 において、両者には“on”と“l'on”との使い分けの比率に関して統計的な有意差があることがわかったが、最終的な言語使用の場面でも、選択される単語に違いがあることが視覚的によくわかる好例であるといえる（図中の網掛け部）。

(直前が et) : [最頻値 : <u>on</u>]	
(直後が con / com) : [頻度 : 20, 65.0%]	<u>l'on</u>
(直後が L) : [頻度 : 48, 100.0%]	on
(それ以外) : [頻度 : 666, 61.1%]	on

図 5 . LeM における“et”と“on / l'on”の選択に関する決定木

図 2・5 は、直後の単語が“con / com”で始まっている場合に限り、“on”と“l'on”の使用が逆転していることを示すものである（図中の波線部）。このような「逆転」は、どの文法書にも該当する記述がなく、今後これをどのよ

うに解釈するべきかの研究を進めなければならない。

5. まとめと展望

今回の調査では、赤間ら(2002)の研究から一步踏み込んで、まず「同一の発音」で終了するケースであっても、単語によって“on”と“l'on”の選択には差があることを示すことができた。このことは、従来の文法書の記述に多大な影響を及ぼしている Vaugelas の解釈や提案だけでは、あらゆる事象を説明しつくすことはできず、「音」という表面だけでなく、「単語そのもの」も考慮に入れて、両者の使い分けの基準を解釈しなければならないということが明らかとなった。

また、「音」と「単語そのもの」を変数として共存させて C5.0 で分析したことで、より精緻な決定木が描けたことはもちろん、同一の単語が直前にあった場合でも、コーパス間で単語選択に差があったり(図 1・2)、従来の文法書では触れられていないような基準で、使い分けられていることが判明したり(図 2~5 中の網掛け部・波線部)したことで、実際の言語表現を悉皆的に調査することの重要性を改めて示すことができたように思う。今後、このような「文法書の記述から逸脱する使い分け」に関しての解釈を進めていくために、音声コーパスを使用した分析というアプローチや、コーパス間で差が出る例があるということで、対象コーパスを国別に設定した、社会言語学的な研究へと発展させることが可能となるだろう。

参考文献

- [1] Paul Dupré, *Encyclopédie du bon français dans l'usage contemporain 2*, Paris, Éditions de Trévise, 1972
- [2] É. Littré, *Dictionnaire de la langue française 2*, Paris, Hachette, 1885
- [3] Claude Favre de Vaugelas, *Remarques sur la langue française*, Geneve, Droz, 1924
- [4] Damourette, L and Pichon, E, *Des mots à la pensée : essai de la grammaire de la langue française Tome IV*, Paris, Collection des linguistes contemporains, 1936
- [5] 赤間啓之, 清水正勝, 清水由美子, 「確率的言語モデルに基づくフランス語の使用例の調査 主語人称代名詞の“on”と“l'on”を例にー」, 情報処理学会・人文科学とコンピュータ研究会, CH56-2(107号), 2002, p.9~16

本論で取り上げなかった、その他の基本論文については、赤間ら(2002)に枚挙してありますので、そちらを参照してください。