

感情・評価を意味する挿入要素の明示されない主体・客体

世 利 彰 規

0. はじめに

0.0. 本研究の目的

本研究は挿入語句のなかでも感情や評価にかかわるものを対象に、そのような挿入語句が誰によって発せられたのか、誰にとってのものなのかを表す要素を明らかにすることを目的とする。例えば、*К счастью, она, моя голубушка, не узнала его.*¹（運のいいことに、あの娘は彼に気づきませんでした）のように、挿入句には誰がその意見を発したのか明示されないことがある。その一方で、*На счастье его, в воротах опять прошло благополучно.*²（彼にとって運のよいことに、門は再び何事もなく通れた）のように、前後に修飾語がついて誰にとってのものなのかははっきり明示されることもある。本研究ではどのような場合に修飾語によってロシア語の挿入要素の表す意味の主体が明示されるのかについて数値化した頻度データから説明する。取り扱う挿入語句は前置詞＋名詞の形をとるものに限定する。

0.1. 本研究の意義

本研究は人文系の言語研究という位置づけでおこなわれた。文法についてだけでなく、文学の文体などに言及することもあるが、基本的にロシア語の言語としての特徴を説明することを目的としている。

本研究の意義に関して次の2点が考えられる。まず、文学作品の解釈に対する貢献が挙げられる。本論考の取り扱う挿入語句・挿入文は文法書において「文法的な一致も支配もされず付加もすることはないため、あたかも文の他の成分と文法的に関係していない」

（якобы не входят в грамматическую связь с остальными членами предложения, так как не согласуются, не управляют, не примыкают）³ や「（文の他の成分との）文法的な従属性が明らかになっていない」（не обнаруживают грамматической зависимости от них (члены

¹ Максим Мошков [http://az.lib.ru/p/pushkin_a_s/text_0430.shtml] (2014年9月1日閲覧).

² Максим Мошков [http://az.lib.ru/d/dostoewskij_f_m/text_0060.shtml] (2014年9月1日閲覧).

³ Руднев А. Г. Синтаксис современного русского языка. М., 1963. С. 201.

этого предложения))⁴と説明され、従来の統語論において異なる性質をもつ要素である
と見なされている。挿入要素の中でも感情や評価を意味するものは発話者の無意識の意見
や感情を表現する。感情・評価を誰が発しているのかがはっきり表されている場合とそう
でない場合を決定する要因を見つけることで、文学作品の登場人物や書き手の無意識的な
意見や感情についての解釈を助けることができる。

第二の意義として考えられるのは、テキストマイニングへの応用である。テキストマイ
ニングとは文章を解析することで必要な情報を抽出することである。テキストマイニング
は、例えば世論の分析や商品開発に用いられている。本論考の対象としているものは評価
や感情を意味する要素である。このような要素を用いてインターネット上に存在する大量
のロシア語の文書に対しテキストマイニングを行なうことでロシア語だけでなく、ロシア
という国の現状を知る手がかりを得ることが期待できる。⁵

0.2. 本研究の特徴

本研究では、ロシア語の文書から必要な情報を抽出するのに計算機の助けを借りている。
解析するテキストはインターネット上のものを使用した。分析に使用するコーパスおよび
言語分析に使用するプログラムは、本研究のため特別に用意された。⁶ したがって、出来
合いのコーパスを利用する場合とは異なり、研究者自身が分析に用いるコーパスそのもの
に関しての知識を持っている。このことは使用するコーパスの長所と短所を認識している

⁴ Валгина Н.С. Синтаксис современного русского языка. М., 1978. С. 249.

⁵ 挿入語句に関連して、テキストマイニング研究において大きな課題となっているのがモダリティ
の問題である。アカデミー70年文法によると文は事実を表す述定 *предикативность* だけではなく、
事実以外の副次的なモダリティの意味を表す要素が存在する。そのような要素は、事実のみを抽出
することをタスクとするテキストマイニングをおこなう場合に邪魔なノイズとなる。このようなノ
イズとしてのモダリティの処理は以下の先行研究で言及されている。(川添愛・齋藤学・片岡喜代子・
戸次大介「確実性判断に関わる意味的文脈アノテーションの試み」『情報処理学会研究報告』NL-189
(12), 2009年, 77-84頁; 荒牧英治・増川佐知子・森田端樹「文書分類と疾患モデルの融合によるソー
シャルメディアからの感染症把握」『自然言語処理』第19巻 第5号, 2012年, 419-435頁。)

⁶ 日本のロシア語学における計量文献学的な先行研究には次のようなものがある：浦井康男「カラ
ムジン『哀れなリーザ』における普通名詞表現の分析：データベースと多変量解析」『スラブ研究』
第42号, 101-116頁, 1995年; 上原順一「MeCabで利用可能なロシア語辞書について」『言語文化研
究』(大阪大学大学院言語文化研究科)第37号, 2011年, 315-322頁。これらの先行研究では、見出
し語の抽出や不特定多数の語の品詞情報の特定や見出し語の決定などを目的としているため、テキ
ストをデータベース化したり、品詞情報を埋め込んだ辞書を作成したりしている。これに対し、本
論考ではテキストファイルから特定の語句を、正規表現から検索するという方針をとった。おおよ
かに言うと、先行研究ではデータを工夫したのに対して本研究ではデータを検索する方法を工夫し
たという点が異なる。また、先行研究の時代と比べてプログラム作成の敷居が低くなっており、自
然言語のテキスト処理に必要なプログラムを作成することが容易になっている。

ということになる。

また、統計処理による数量データによって結論を導いていることも特徴の一つである。

本研究は言語データの収集に計算機を用いているが、言語の分析そのものは人手で行なっていることを強調しておきたい。本研究では計算機を、自動化できる作業の効率化のために使用しているにすぎない。本質的な部分は従来の人文系の言語研究と変わらない。

0.3. 概要

本稿の概要は次のようになる。まず取り扱う問題を提起する。そして挿入語句や主体・客体を表す語句は文法的にどのような分野に属するかについて先行研究を概観し、作業仮説を提起する。つぎに分析に使用する言語データはどのようなものなのか、どのように入手したのかについて説明する。そして入手したデータについてどのような分析をおこなったのか、分析手法について明らかにする。その後、分析した結果、提起した仮説は有意なものであったのかどうかを述べる。最後に、結論と本研究でおこなうことのできなかったことについて触れる。

1. 問題提起

本稿で取り扱う問題は次のようなものである：ロシア語の挿入語句 *вводные слова* はいくつかに分類され、その中には感情や評価を表すものが存在する。上で述べたように挿入語句は文から独立した成分と見なされるため、「誰が」や「誰に対して」そのような感情なり評価なりをもったのかについての情報がはっきり表示されない場合がある。例えば：

И он стал на колени среди тротуара, *к счастью*, на этот раз пустынного.⁷

そして彼は歩道の上に膝まずいたが、幸いなことに、そのとき通りには誰もいなかった。

К счастью, в стороне блеснул тусклый свет и помог мне найти другое отверстие наподобие двери.⁸

運のいいことに、脇の方におぼろげな光がちらついていたので私は別の戸口のような隙間を見つけることができた。

一方で、誰がそのような感情を持ったのか、修飾語によって明らかにしている場合も存在

⁷ Максим Мошков [http://az.lib.ru/d/dostoewskij_f_m/text_0060.shtml] (2014年9月1日閲覧).

⁸ Максим Мошков [http://az.lib.ru/l/lermontow_m_j/text_0410.shtml] (2014年9月1日閲覧).

する：

Кити танцевала в первой паре, и, *к ее счастью*, ей не надо было говорить, потому что Корсунский все время бегал, распоряжаясь по своему хозяйству.⁹

キティは最初の組で踊ったが、彼女にとって幸運なことに、彼女は話す必要はなかった。というのも、コルスンスキイが常に場を切り盛りしながら走り回っていたからだった。

Страшная гроза застала меня на дороге. *К счастью*, невдалеке виднелось село.¹⁰

道で私は恐ろしい雷雨に見舞われました。幸運なことに、さほど遠くないところに村が見えました。

このような例は 19 世紀文学作品だけではなく、インターネット新聞のような報道のジャンルにも見られる：

У него наступила остановка дыхания, но, *к счастью*, прямо на поле медикам удалось реанимировать Вадима.¹¹

彼の心肺は停止したが、幸運なことに、すぐに芝生の上で医師がワジームを蘇生した。

В доме, где жила семья, возник пожар. В результате женщина решила бросать малышей из окна. *К счастью*, их поймали соседи.¹²

家族が住んでいた家で火事が起こった。結果として女性は子どもたちを窓から投げ下ろすことに決めた。幸運なことに、隣人が彼らを受け止めたのだった。

Канадец случайно выстрелил в себя из винтовки во время охоты на мышь, сообщает Orange News. В результате инцидента 40-летний Дейл Уитмелл не получил серьезных повреждений — к счастью для него, пуля лишь едва задела кожу на его лбу.¹³

あるカナダ人がねずみ取りをしている間に偶然にもライフル銃で自分を撃ってしまったことを Orange News が伝えた。この事故の結果 40 歳のデイル・ウイトメルが受けた傷はそれほどひどいものではなかった。彼にとって幸運なことに、弾丸は彼の額の皮膚をかすめただけだ

⁹ Максим Мошков [http://az.lib.ru/t/tolstoj_lew_nikolaewich/text_0080.shtml] (2014 年 9 月 1 日閲覧).

¹⁰ Максим Мошков [http://az.lib.ru/t/turgenew_i_s/text_0081.shtml] (2014 年 9 月 1 日閲覧).

¹¹ Первый канал [<http://www.1tv.ru/news/polit/264596>] (2014 年 9 月 1 日閲覧).

¹² Первый канал [<http://www.1tv.ru/news/polit/258233>] (2014 年 9 月 1 日閲覧).

¹³ Лента [<http://txt.ru/news/2012/07/27/mouse/>] (2014 年 9 月 1 日閲覧).

った。

Возгорание в старинном театре произошло за 15 минут до детского спектакля “Вредные советы”, который должен был начаться в 12:00. К счастью всех участников и зрителей, всего порядка 350 человек, успели вывести на улицу.¹⁴

古い劇場が発火したのは 12 時に始まる予定であった児童劇「有害な忠告」の開演 15 分前であった。出演者全員と 350 人におよぶ観客にとって幸運なことに、通りに避難する時間があった。

前置詞＋名詞の形の感情・評価を意味する挿入語句（特に по＋与格、к＋与格の形。以降の箇所では「感情・評価を意味する挿入語句」と呼ぶ）を対象に、「誰が」「誰・何に対し」（以降の箇所では「主体・客体」と呼ぶ）発したかを表示される場合と明示されない場合の規則は何か、という問題を本研究では取り扱う。本研究で提起される問題は「感情・評価を意味する挿入語句において主体・客体を表す語句が出現することに関わる要因はどのようなものであるか」とする。

2. 先行研究

アカデミー60年文法は本研究で取り扱う挿入語句について次のような説明をしている：

§1222. 挿入語と挿入句は文の中で伝えられていることについての感情的な評価を表現しうる。この主の語句は喜びや満足、悲しみ、悔恨、驚き、憤り、困惑等をあらわす。¹⁵

アカデミー60年文法において挿入語句とモダリティを表すモーダルな語句との関連性は指摘されていない。

次の文献の記述は前置詞 к を伴った与格による感情表現とモーダルな語句とをはっきり区別している：

この種のモーダルな語 модальные слова は、挿入されたシンタグムの生産的なタイプに意味において類似している。このようなシンタグムは、感情を意味したり、そのような感情を呼び起こすような状況を意味したりする名詞の与格形を伴った前置 к から形成される。：к

¹⁴ Правда [http://www.pravda.ru/news/accidents/03-11-2013/1180674-shkola-0/] (2014 年 9 月 1 日閲覧)。

¹⁵ Грамматика русского языка. М., 1954. Т. 2. Ч. 1. С. 153.

сожалению, к несчастью, к изумлению, к счастью, к общему восхищению など。¹⁶

また **вводные слова** (挿入語) と **вставные конструкции** (入れ子構造) は似ているが、異なるものである。入れ子構造は括弧やダッシュなどの記号に括られたものと考えてよい。文法書で区別について特に言及されない場合もあるが、ローゼンターリは **вставные конструкции** に関して、次のように定義している。¹⁷

入れ子構造の文体的機能は、語彙的構成要素 (慣用句となる段階にある書き言葉や話し言葉の語彙) や、統語的な構造 (語結合の特徴や文のタイプ)、感情的・表出的なニュアンスによって条件づけられる：

Итак, я соглашаюсь с большинством голосов, которое решило, что всего благоразумнее и безопаснее внутри города ожидать осады, а нападения неприятеля силой артиллерии и *(буде окажется возможным)* вылазками — отражать.

したがって、わたしは、町の中で包囲されるのを待って、敵の攻撃に対しては砲兵隊の戦力と (もし可能のようなら) 奇襲攻撃とによって反撃することが何より賢明で安全である、と決定した大多数の意見に賛成の意を表します。

Литвинов остался на дорожке; между им и Татьяной — *или это ему только чудилось?* — совершалось что-то... бессознательно и постепенно.

リトヴィノフは道に取り残された；彼とタチヤーナの間には—それともこれは彼の単なる思い過ごしなのだろうか？—何かが無意識にそして次第に起こっていた。

上の例文から入れ子構造 (**вставные конструкции**) は、括弧あるいはダッシュに囲まれたフレーズを指していると考えられる。今回扱うのはそれとは異なる挿入語 (**вводные слова**) とする。

3. 作業仮説の構築

本研究で考察対象とする、感情・評価を意味する挿入語句に主体・客体を表す修飾語が付くことに関わってくる要因は、いくつかが考えられる。そのため、本論考では上で提起した問題に対し、一つだけでなく 3 つの作業仮説を構築する。個々の仮説に対してデータを収集し統計解析を行なって、仮説の有意性について検討する。

¹⁶ Виноградов В.В. Русский язык. М., 1972. С. 578

¹⁷ Розенталь Д.Э. Справочник по русскому языку. М., 2008. С. 351-352.

3.1. 語彙に関わる要因

まず考えられるのが語彙的な要因である。つまり、どのような意味を持つ名詞が挿入句を構成しているのかが挿入語句に主体を明示する要素が付加されるか否かを決定するというのである。

単純に考えて、挿入語句の表す意味によって「誰にとって」「誰の」を明示する頻度は違ってくると思われる。たとえば「к счастью (幸運なことに)」「к сожалению (残念ながら)」といった語句は「誰にとって」幸運か、残念かを必ずしも明らかにしなくても成立することは多い。その一方で、「к радости (喜ばしいことに)」「к удивлению (驚いたことに)」のような挿入語句は誰が喜んだか、誰が驚いたかを明らかにしないと言葉が足りないと感じられる場合がある。以上のことから、挿入語句を構成する名詞が主体・客体がつく頻度に影響を与える可能性は十分考えられる。

また、この作業仮説の根拠として、前置詞＋名詞からなる「情報源」を表す по словам, по сообщению などの挿入語句には必ず его や моему など発信源が明示されることが挙げられる。これは挿入表現を構成する名詞による違いからくるものであると考えられる。このことも挿入語句の主体・客体の明示が挿入語句を構成する名詞に関係するのではないかという根拠になると考えられる。

これらの挿入語句は「情報源」を意味するので、「誰が」を意味する主体を表す語が必ず付く。一方で、本研究の対象とする挿入語句の意味する「評価」や「感情」の意味も「誰が」「何に対して」を表すことが可能である。しかし、実際にはそれらを意味する語は必ずしも存在する訳ではない。

60 年と 80 年のアカデミー文法においては、感情・評価を意味する挿入語句として次のものが挙げられている。

60 年文法の分類

60 年文法において「文の中で述べられていることに関する感情的な評価を表わす挿入語句」として分類が行なわれている。次のような語句がリストとして挙げられる：

к счастью, к несчастью, к радости, к удивлению, к изумлению, к сожалению, к стыду, к огорчению, к прискорбию, к досаде; на счастье, на радость, на беду, на ту беду; по счастью, по несчастью; странное дело, удивительное дело, известное дело, неровен час, воля твоя (ваша), грешным делом, некоторым образом, чего доброго, нечего греха таить; как на беду, как нарочно;

как хотите, если хотите.¹⁸

「感情的な評価」というカテゴリーに挿入語句が分類されたが、本研究で問題とする主体・客体についてははっきりと言及はされていない。文例において *к моей великой досаде* のような例が挙げられているのみである。

80 年文法の分類

80 年文法においては「発話者自身の主観的態度、感情的な反応、理性的な評価を表わす役割をもつ挿入語句」として分類されている。次のようなものが例としてあげられている：

発話者自身による主観的態度、感情的反応、知情的評価。喜び、賞賛、肯定的評価：*на счастье, к счастью, к радости, к удовольствию кого-н., что хорошо, что еще лучше*; 哀れみ、不承認、否定的評価：*к несчастью, к сожалению, к стыду чьему-н., к неудовольствию, к огорчению кого-н., к прискорбию чьему-н., к досаде, на беду, по несчастью, грешным делом, как на беду, как нарочно, что еще хуже, что обидно, увы*; 驚き、困惑：*к удивлению кого-н., к изумлению, странно, странное дело, непонятное дело, удивительно, удивительное дело*;¹⁹

60 年文法と比べ、新たに本研究で取り扱う主体・客体を表す表現についての項目が例えば、*к удовольствию кого-н.* や *к прискорбию чьему-н.* のような形で記載されている。また、個々の感情について細分化が行なわれている。

以上のことから、本論考で取りあつかう評価や感情を意味する挿入語句を構成する名詞の意味によって、挿入語句に修飾語が付くか否かが決まるのではないかという作業仮説を立てる。この仮説の検証には統計的仮説検定ではなく、データによる分類が必要になる。

3.2. 他の種類の修飾語の付随

この作業仮説は実際に人手で言語分析を行なっているときに感じた直感によるものである。詳しくは後述するが、人手によって挿入語句を「主体・客体」を表す語句のついてあるものについていないものへと分類して分析をおこなった。その際、特に文学作品において、「主体・客体」を表す語句「多いに *великий*」や「少なからずの *немалый*」といった性質形容詞が挿入語句に多く付随して現れるような印象を受けた。

¹⁸ Грамматика русского языка. М., 1954. Т. 2. Ч. 1. С. 153.

¹⁹ Русская Грамматика. М., 1980. Т. 2. С. 229.

«Но, к моему большому сожалению, не всем мечтам суждено сбываться. По этой причине я приняла решение о завершении спортивной карьеры в биатлоне», — отметила Юрьева.²⁰

「でも、私にとってとても残念なことなのですが、すべての夢が叶うわけではなかった。だから私はバイアスロンの現役スポーツ選手としての自分のキャリアを終わらせることに決めました」とユーリエヴァはコメントした。

“К нашему большому сожалению, весь европейский тур группы Motorhead отменен. Концерт в Москве не стал исключением”, — сообщает агентство.²¹

「私たちにとってはとても残念なことなのですが、グループ Moterhead のヨーロッパツアーの全日程は中止になりました。モスクワでのコンサートもその例外ではありません」と代理人は伝えた。

После чего крокодил, к немалому удивлению зрителей, ретировался. Посетители проводили бесстрашную кошку аплодисментами.²²

その後、見物人たちにとって少なからず残念なことであったが、クロコダイルは退却していた。客たちは勇敢な子猫に喝采を送った。

反対に主体・客体を表す表現がない場合、挿入語句を構成する名詞に修飾語が付くことも少ないように思われる。これはインターネット新聞で多い印象がある：

К счастью, большинство этих посылок вовремя были перехвачены, поэтому пострадал только один человек — развозивший почту курьер, у которого в руках взорвался пакет, адресованный мексиканскому посольству в Афинах.²³

幸運なことにこれらの郵便物は折よく取り押さえられたため、被害を受けたのは郵便局へ荷物を運んでいた配達人ただ一人であった。アテネにあるメキシコ大使館宛ての小包は彼の手の中で爆発した。

“Считаю, что это в любом случае полезно, потому что мы пытаемся соединять подходы и, может быть, устранять причины для эскалации насилия. Оно, к сожалению, продолжает

²⁰ Лента [<http://txt.ru/news/2014/02/27/theend/>] (2014 年 9 月 1 日閲覧).

²¹ Правда [<http://www.pravda.ru/news/world/02-07-2013/1163876-european-0/>] (2014 年 9 月 1 日閲覧).

²² Первый канал [<http://www.1tv.ru/news/polit/227216/>] (2014 年 9 月 1 日閲覧).

²³ Лента [<http://txt.ru/news/2011/07/19/bombers/>] (2014 年 9 月 1 日閲覧).

использоваться. И чем быстрее будут прекращены военные действия, лучше будет для всех, кто живет в Ливии”, — подчеркнул президент РФ.²⁴

「私はこれをいかなる場合においても有益であると思っている。なぜなら私たちはアプローチを総合しようと試みており、そして、多分、暴動がこれ以上エスカレートする要因を取り除こうとしているからだ。暴動は、残念なことに、利用され続けている。だから軍事行動がなるべく早く終結することがリビアの住民にとって望ましいことになるだろう」とロシアの大統領は強調した。

Молодой человек отправился покататься на лодке и заплыл достаточно далеко от берега. Он был одет лишь в футболку и шорты. К счастью, юноша догадался захватить с собой мобильный телефон.²⁵

青年はボートに乗りに出かけ、岸から遠くに流されてしまった。彼は運動シャツと短ズボンしか着ていなかった。幸運なことに、青年は携帯電話を持っていたことに気づいた。

つまり、単に к удивлению と簡単にコメントするかあるいは к моему великому удивлению と程度や誰の感情なのかという発言の主体を付け加えて発せられることが多く、к моему удивлению とか к великому удивлению という表現が出てくことは少ない。

ここで「他の種類」という表現を用いているが、本研究で取り扱っている主体・客体を表す表現の一部は物主形容詞や物主形容詞的代名詞であるためである。アカデミー80年文法でこれは関係形容詞と性質形容詞の2つに形容詞は大分されている。²⁶

§1295. 性質形容詞は対象そのものやそれによって明らかにされるものに付属する性質を意味する。しばしばそれらは強さのさまざまな度合いを特徴付けることができる。[...] これには色彩、空間、時間、物理的そしてその他の質的・量的・精神気質といった性質や特質をもった語が分類される。громкий, густой, тяжелый, малый, важный など

§1296. 関係形容詞は事象や他の特徴との関係を通して意味づけを行なう。[...] 性質形容詞は動機づけられていた語や動機づけられていない語によって提示される。それとはちがって関係形容詞は他の別の品詞すなわち名詞 (отцов, сестрин, комсомольский [...]) や動詞 (танцевательный [...]), 数詞 (четвертый, девятый [...]) や副詞 (ближний, прежний [...]) に属する語によって動機づけされる。

²⁴ Лента [<http://lenta.ru/news/2011/05/27/medvedev/>] (2014年9月1日閲覧).

²⁵ Правда [<http://www.pravda.ru/news/world/europe/european/14-07-2014/12160480/>] (2014年9月1日閲覧).

²⁶ Русская Грамматика. М., 1980. Т. 1. С. 541-542.

関係形容詞はさらに 1) 本来の関係を表すもの (собственно относительные), 2) 順序を表すもの (порядковые), 3) 代名詞に特有な意味をもつもの (местоименные) の 3 つのカテゴリに下位分類される。この第三の「代名詞に特有な意味をもつもの」のうち物主代名詞に分類されるものは、本研究で問題とする主体・客体を表す挿入語句の要素の一部をなしている。

感情や評価といった意味は「強さのさまざまな度合い」によって形容されることは明らかであり、主体・客体を表す表現とは異なるものに分類される。また всеобщий や собственный といった物主代名詞以外の関係形容詞も「他の種類」に含める。

人手による分析をおこなっていて感じた直感的な印象「主体・客体を表す語句の出現は他の種類の修飾語が付くかどうかに影響する」という仮説を立てる。

3.3. ジャンルに関わる要因

最後に構築する作業仮説として文体的な要因がある。すなわち、テキスト全体の文脈が挿入語句の主体が明示されるか否かに作用するのではないかというものである。

報道のジャンルにおいて必ず「誰が」その意見や情報を発したかが明らかにされる「по словам」「по данным」といった挿入語句が多く用いられる。このような挿入語句は「情報の発信元を表す」挿入語句と呼ばれる。²⁷ このような語句がインターネット新聞という報道のジャンルで多く用いられるのは「報道」という文体のジャンルに関係していると考えられる。発信元をはっきりさせることと同様に、感情や評価を表す場合に「誰が」そのような感情や評価を抱いたのかを報道のジャンルでは明らかにすることが多いのではないだろうか。以上が感情・評価を意味する挿入語句における主体・客体を表す語句の出現にテキストのジャンルが関係しているのではないかと考える理由である。

以上のことをふまえて、文学作品とインターネット新聞という文体の差が、評価・感情をもつ主体がはっきり表示されるか否かに関わっているという作業仮説が立てられる。

4. データの収集

4.1. Web スクレイピング

本研究では Web スクレイピングの手法を用いて収集したロシア語の言語データを分析に使用している。Web スクレイピングとはプログラムやソフトウェアを使用してウェブ上

²⁷ Грамматика русского языка. Т. 2. Ч. 1, М., 1954. С. 154.

の電子テキストを自動的に取得することである。本研究においては、Web スクレイピングをおこなってテキストを取得した（分析に使用したプログラムの例は付録として掲載してある）。収集したテキストは一意となるように年月日と番号を振って保存した。

4.2. インターネット新聞

上記の手続きによって取得した Web リソースについて一覧表にして提示する：

新聞名	URL	文書数	平均語数	語数最頻値
Лента	http://txt.ru	114,210	203.0	195
Независимая газета	http://www.ng.ru	38,145	118.2	99
Правда	http://www.pravda.ru	68,567	176.0	144
Первый канал	http://www.1tv.ru	44,912	253.1	140

上の表の「URL」とはそのテキストの取得先のトップページのアドレス、「文書数」とはテキストファイルの数、「平均語数」とはそれぞれの新聞の語の総和を文書数で割った算術平均、「語数最頻値」はそれぞれの新聞リソースにおいてある 1 テキストが何語で構成されることが最も多いのかを指す。以上の合計 265,834 文書を分析に使用する。

4.3. 電子テキスト化された文学作品や評論

上で挙げたインターネット新聞に加えて、文学テキストも本研究における分析で用いることにする。19 世紀ロシア文学者によって書かれた文書全般（作品に限らない）を対象とする。Библиотека Максима Мошкова (<http://www.lib.ru/>)²⁸ から Web スクレイピングで取得したテキストを分析に使用する。使用した文学作品名を作家ごとに一覧表として表示する：

作者名	文書数
プーシキン	1.Сказки

²⁸ この電子テキストを利用している先行研究として水野晶子・藤村逸子「身体部位名詞を伴う再帰構文における格の問題—ロシア語とフランス語の比較」『国際開発研究フォーラム』（名古屋大学大学院国際開発研究科）第 26 号，2004 年，65-87 頁，がある。ここで考慮しておかなくてはならないことは、ソ連時代の正書法を考慮する必要があるということである。正書法は句読点の打ち方に関係し、これは 19 世紀という時代的なものと個々の作家のくせに由来する。ソ連時代には 19 世紀に執筆した文学者のつけた句読点はソ連時代の正書法規則に合わせて書き換えてあり、19 世紀のロシア語の特徴や個々の作家の文体的な特徴といった要素は捨象されることになった。一方で、本研究は文学作品の文体や通時的な特徴ではなく、ロシア語そのものの特徴に関心を向けている。したがって、個々の文学者による文書を均等な言語データとして扱うこととし、この問題について考慮しない。そのため、インターネット上の文書をそのまま使用する。

	2.Евгений Онегин 3.Арап Петра Великого 4.Повести покойного Ивана Петрович Белкина 5.Рославлев 6.Дубровский 7.Пиковая дама 8.Кирджали, Капитанская дочка 9.Борис Годунов
ゴーゴリ	1.Вечера на хуторе близ Диканьки, часть первая 2.Вечера на хуторе близ Диканьки, часть вторая 3.Старосветские помещики 4.Тарас Бульба 5.Вий 6.Повесть о том, как поссорился Иван Иванович с Иваном Никифоровичем 7.Невский проспект 8.Нос 9.Портрет 10.Шинель 11.Записки сумасшедшего 12.Мертвые души. Том первый. 13.Мертвые души. Том второй. 14.Выбранные места из переписки с друзьями
レールモンτροφ	1.Герой нашего времени 2.Кавказец 3.Вадим 4.Панорама Москвы 5.Княгиня Лиговская
トゥルゲーネフ	1.Рудин 2.Дворянское гнездо 3.Накануне 4.Отцы и дети 5.Дым

	6.Новь
	7.Жид
	8.Бретёр
	9.Петушков
	10.Дневник лишнего человека
	11.Муму
	12.Записки охотника
	13.Незавершенное
	14.Яков Пасынков
	15.Фауст
	16.Затишье
	17.Поездка в Полесье
	18.Ася
	19.Первая любовь
	20.Несчастливая
	21.Степной король Лир
	22.Вешние воды
	23.Часы
	24.Песнь торжествующей любви
	25.После смерти (Клара Милич)
	26.Прозаические наброски
	27.Капля жизни
	28.Собственная господская контора
	29.Переписка
	30.Отрывки из воспоминаний - своих и чужих
	31.Senilia
	32.Повести и рассказы (Варианты)
	33.Андрей Колосов
	34.Три портрета
	35.Три встречи
	36.Постоялый двор
	37.Два приятеля
	38.Степан Семенович Дубков и мои с ним разговоры

	<p>39.Призраки</p> <p>40.Довольно</p> <p>41.Довольно</p> <p>42.История лейтенанта Ергунова</p> <p>43.Странная история</p>
ドストエフスキー	<p>1.Униженные и оскорбленные</p> <p>2.Игрок</p> <p>3.Преступление и наказание</p> <p>4.Бесы</p> <p>5.Братья Карамазовы. Часть 1</p> <p>6.Братья Карамазовы. Часть 2</p> <p>7.Братья Карамазовы. Часть 3.</p> <p>8.Братья Карамазовы. Часть 4.</p> <p>9.Двойник</p> <p>10.Роман в девяти письмах</p> <p>11.Господин Прохарчин</p> <p>12.Хозяйка</p> <p>13.Ползунков</p> <p>14.Чужая жена и муж под кроватью</p> <p>15.Честный вор</p> <p>16.Неточка Незванова</p> <p>17.Маленький герой</p> <p>18.Дядюшкин сон</p> <p>19.Село Степанчиково и его обитатели</p> <p>20.Скверный анекдот</p> <p>21.Записки из подполья</p> <p>22.Крокодил</p> <p>23.Вечный муж</p> <p>24.Бобок</p> <p>25.Сон смешного человека</p> <p>26.Петербургские сновидения в стихах и в прозе</p> <p>27.Два самоубийства</p> <p>28.Кроткая</p>

トルストイ	<ol style="list-style-type: none"> 1.Семейное счастье 2.Война и мир. Том 1 3.Война и мир. Том 2 4.Война и мир. Том 3 5.Война и мир. Том 4 6.Война и мир. Первый вариант романа 7.Анна Каренина 8.Воскресение 9.Детство 10.Отрочество 11.Юность 12.Два гусара 13.Поликушка 14.Казаки 15.Холстомер 16.Смерть Ивана Ильича 17.Дьявол 18.Крейцера соната 19.Отец Сергей 20.Хаджи-Мурат 21.Севастополь в декабре месяце 22.Севастополь в августе 1855 года 23.Альберт 24.Набег. Рассказ волонтера 25.Рубка леса. Рассказ юнкера 26.Метель 27.Разжалованный (Из кавказских воспоминаний) 28.Три смерти 29.Хозяин и работник 30.Алеша Горшок 31.Ягоды 32.Корней Васильев 33.Правила для педагогических курсов и заметки на тему народного
-------	---

	<p>образования</p> <p>34.Первая ступень</p> <p>35.Два письма к М. Ганди</p> <p>36.Ответ польской женщине</p> <p>37.Благо любви</p> <p>38.Время пришло</p> <p>39.К духовенству</p> <p>40.Собрание малоизвестных философских, религиозных и публицистических произведений из 17-го тома Псс</p> <p>41.По поводу заключения В.А.Молочникова</p> <p>42.О душе и жизни ее вне известной и понятной нам жизни</p> <p>43.О значении христианской религии</p> <p>44.Патриотизм или мир?</p> <p>45.Carthago Delenda Est</p> <p>46.Доклад, приготовленный для конгресса о мире в Стокгольме</p> <p>47.О безумии</p> <p>48.По поводу конгресса о мире (Письмо шведам)</p> <p>49.Письмо студенту о праве</p> <p>50.Неделание</p> <p>51.Богу или мамоне?</p> <p>52.Письмо революционеру</p> <p>53.Удивительные существа</p> <p>54.Предисловие к статье Эдуарда Карпентера "Современная наука"</p> <p>55.Студенческое движение 1899 года</p> <p>56.Патриотизм и правительство</p> <p>57.Исследование догматического богословия</p> <p>58.В чем моя вера?</p> <p>59.Исповедь</p> <p>60.Религия и нравственность</p> <p>61.Христианское учение</p> <p>62.Греческий учитель Сократ</p> <p>63.Определение Святейшего Синода</p> <p>64.Ответ Синоду</p>
--	--

	65.Соединение и перевод четырех Евангелий 66.Учение Христа, изложенное для детей 67.Путь жизни 68.Не могу молчать (1-я редакция) 69.Что такое религия и в чем сущность ее 70.Церковь и государство 71.Славянскому съезду 72.Китайскому народу от христианина 73.Воззвание 74.Послесловие к книге Е.И. Попова "Жизнь и смерть Евдокима Никитича Дрожжина. 1866-1894" 75.Как читать евангелие и в чем его сущность? 76.Отрицает Иисуса как Искупителя 77.Против 7-ми таинств 78.Против обожествления Иисуса 79.Против троицы 80.Критика православного богослужения 81.Не могу молчать 82.Закон насилия и закон любви
--	---

上で挙げた文書内に含まれる語数に関して算出した基本等計量を一覧表にして表示する。これらの統計量はテキストについての大体の傾向を示すためのものである：

作者名	文書数	全語数
プーシキン	10	144,199
ゴーゴリ	14	357,087
レールモントフ	4	89,030
トゥルゲーネフ	44	840,301
ドストエフスキー	28	1,225,078
トルストイ	82	2,206,087

以上 19 世紀の文学者 6 人の手による計 182 文書を言語データに加えて分析に使用する。

5. 言語分析

5.1. 計算機による検索

取り出す表現は以下の 16 の挿入語句とする。

1.к счастью	2.к изумлению
3.на счастье	4.к удивлению
5.по счастью	6.к радости
7.к несчастью	8.на радость
9.на несчастье	10.к сожалению
11.по несчастью	12.к прискорбию
13.к удовольствию	14.к досаде
15.к неудовольствию	16.к огорчению

言語分析の手続きの一部は人手によって処理や分析を行なう。テキストから形だけで特定できる挿入語句を取り出し、数え上げることは機械的な作業である。それに対し、取り出されたリストから不適切なものや明らかな重複を除外したり、それらの挿入語句の意味や文脈を加味して考察を行なったりするよう計算機に指示することは難しい。集められた言語データを解釈することはむしろ人手で行なった方が、容易に正確な結果を得ることができる。以上の理由から集められた挿入語句の分析は人手で行なう。

5.2. 人手による分析の流れ

まず分類するのに戸惑う例についてどのように扱ったのか説明する。まず、前に **и**, **но**, **еще** などの語や後ろに **еще** や **же** といった語が付いていても、コンマで区切られていれば挿入語句の一つとして数える：

*И к удивлению многих, мусульманская молодежь не осталась в стороне. Дети из арабских и еврейских семей здесь часто играют в одной команде.*²⁹

そして多くの人が驚くことに、ムスリムの若者は一方の側にとどまることをしなかった。アラブ家庭の子どもたちもユダヤ人家庭の子どもたちもここではしばしば同じ一つのチームで競技している。

²⁹ Первый канал [<http://www.1tv.ru/news/polit/256014>] (2014 年 9 月 1 日閲覧).

В помещении также находилась девушка, на которую преступник также навел оружие и дважды нажал курок. *Однако к счастью свидетельницы*, пистолет дал осечку.³⁰

建物の中には少女もあり、犯人は彼女に銃器を向けて二度引き金を引いた。しかし証人となった彼女にとって幸運なことに、発砲は不発に終わった。

К несчастью же, весьма часто можно найти, особенно в посланиях Павла, такое выражение учения, которое может вовлекать читающих в недоразумения, затемняющие самое учение.³¹

不運なことに、とりわけパウロの書簡には、そのために読む人にとって教えそのものが理解しにくくなり、誤解されうるような表現がはなはだ頻繁に散見される。

二つ名詞がある場合、たとえば *к несчастью и удивлению своему* のようなものは最初の名詞によって構成される挿入語句一つ分と見なして数える（つまり *к несчастью и удивлению своему* は *к несчастью* 一つ分として数える）。

Он пощупал пульс и, *к удивлению и неудовольствию своему*, заметил, что пульс был лучше.³²

彼は脈をとった。そして彼にとって驚くべき、そしていまいましいことに、脈がよりよくなっていることに気づいたのだった。

Наташа находилась в этом состоянии столбняка с нынешнего утра, с того самого времени, как Соня, *к удивлению и досаде графини*, непонятно для чего, нашла нужным объявить Наташе о ране князя Андрея и о его присутствии с ними в поезде.³³

今朝ソーニャは、公爵婦人が驚き苛立ったことに、何のためか分からないが必要を感じてナターシャにアンドレイ公爵の負傷と彼が同じ汽車に乗っていることを知らせてしまい、そのとき以来ナターシャは放心状態にあった。

言語データから必要なデータの抽出は以下のような手順を踏んで人手によって行なわれる。

- 1.機械によって取得したテキストを可能な限り編集する（余分な部分を取り除く）。
- 2.重複したものを除外する。
- 3.不適切な例を除外する。

³⁰ Правда [http://www.pravda.ru/news/accidents/14-10-2013/1178076-murder-0/] (2014 年 9 月 1 日閲覧)。

³¹ Максим Мошков [http://az.lib.ru/t/tolstoj_lew_nikolaewich/text_0510.shtml] (2014 年 9 月 1 日閲覧)。

³² Максим Мошков [http://az.lib.ru/t/tolstoj_lew_nikolaewich/text_0060.shtml] (2014 年 9 月 1 日閲覧)。

³³ Максим Мошков [http://az.lib.ru/t/tolstoj_lew_nikolaewich/text_0060.shtml] (2014 年 9 月 1 日閲覧)。

- 4.主体・客体を表す語句が付いたものを見つけ、必要なデータを記録する。
- 5.特異な例を記録する。

この中でも特に手順4について詳しく説明する。手順2は必要なデータを記録する作業である：

- 1.主体・客体を表す語句が付いた例を数え上げる。
- 2.作業仮説2で取り扱う主体・客体を表す語句に相応する語句がテキストの前の部分で出現したか否かを調べる。
- 3.その他の興味深い例文を記録する。

以上のように説明した手順によって言語情報を数値によるデータとして統計処理しやすいように記録する。

5.3. 分析結果

まず人手による言語分析から明らかになったデータを表形式にして提示する。感情・評価を意味する挿入語句に付随する主体・客体を表す語句はどのようなものがあるのかについてのデータを表形式にして挙げる：

挿入語句	インターネット新聞	19世紀文学者の文書
к счастью	кого-чего, для кого-чего,	кого-чего, для кого-чего, моему
на счастье	кого-чего, мое	кого-чего, мое
по счастью	кого-чего, мое	
к несчастью	кого-чего, для кого-чего,	кого-чего, вашему, своему
на несчастье		мое
по несчастью		
к удовольствию		кого-чего, своему, моему
к неудовольствию	кого-чего	
к изумлению	кого-чего	кого-чего, моему, своему,
к удивлению	кого-чего, для кого-чего, моему, нашему, своему	кого-чего,
к радости	кого-чего	кого-чего, нашей, моей, своей,
на радость	кого-чего, кому-чему,	кого-чего, моему, своему
к сожалению	кого-чего, моему, нашему, своему, для кого-чего, к	кого-чего, моему

	кому-чему	
к прискорбию		кого-чего,
к досаде		кого-чего, моей, своей
к огорчению	нашему	кого-чего, своему

以上にあげた物主代名詞と名詞斜格形、前置詞＋名詞斜格を本研究で取り扱う「主体・客体を表す語句」として定義する。

文法書には載っていない主体・客体の形を見つけることができた。具体的には *к счастью для кого-чего, на радость кому-чему, к сожалению для кого-чего/кому-чему* などが挙げられる：

*К счастью для Гленна Форда, за это время приговор не привели в исполнение, и нашелся новый свидетель, чьи показания в корне изменили всё.*³⁴

グレン・フォードにとっては幸運なことに、この間に判決は執行に移されることはなく、新たな証人が見つかって彼の証言はすべてを根底から覆すこととなった。

*Иностранцев традиционно много в Пекине, на радость местным таксистам, к досаде других участников дорожного движения.*³⁵

現地のタクシー運転手には喜ばしいことに、他の運転手たちにはいまいましいことに、伝統的に北京には外国人が多い。

*Стадионы сегодня стали территорией беззакония, бескультурия, и, к сожалению к большому, в первых рядах получается и руководитель Чечни.*³⁶

今日スタジアムは無法の有史以前の地帯と化し、多くの人にとって残念なことに、最前列にいたのがチェチェンの指導者だった。

*К сожалению для немцев, они не очень хорошо себя под Прохоровкой показали, и к счастью для нас.*³⁷

ドイツ人にとっては残念なことに、かれらはプロハロフカの近くでうまく本領を発揮することができなかったが、これは私たちにとっては幸運なことであった。

³⁴ Первый канал [http://www.1tv.ru/news/polit/253991] (2014 年 9 月 1 日閲覧).

³⁵ Первый канал [http://www.1tv.ru/news/polit/223736] (2014 年 9 月 1 日閲覧).

³⁶ Первый канал [http://www.1tv.ru/news/polit/228665] (2014 年 9 月 1 日閲覧).

³⁷ Первый канал [http://www.1tv.ru/news/polit/237429] (2014 年 9 月 1 日閲覧).

これらはいずれもインターネット上の例である。アカデミー文法書と違った形が見られる理由について様々な理由が考えられるが、この問題についてここではこれ以上取り扱わない。

5.4. その他の副次的なデータ

本稿の主問題とは別に、分析によって明らかになった記述的なデータを表の形式で提示する：

挿入語句	インターネット新聞	19 世紀文学者の文書
к счастью		
на счастье		
по счастью		
к несчастью		
на несчастью		
по несчастью		
к удовольствию		необыкновенному величайшему неописанному
к неудовольствию		великому
к изумлению	огромному	величайшему неописанному общему собственному великому немалому крайнему горестному великому
к удивлению	огромному	чрезвычайному несказанному немалому крайнему всеобщему величайшему собственному пущему великому истинному большему
к радости	огромному	величайшей неописанной немалой великой
на радость	огромной	
к сожалению	величайшему совеликому огромному глубокому	искреннему общему

	большому	
к прискорбию		величайшему
к досаде		великой величайшей несказанной страшной неописанной
к огорчению		немалому великому

まず目につくのはインターネット上のコーパスにおいて主体・客体以外の修飾語は **огромный** が用いられることが多いということである。к изумлению, к удивлению, к радости, на радость にはもっぱら **огромный** が用いられている。それに対し、今回調査した 19 世紀文学においては様々な修飾語が付与される。

この節では感情・評価を意味する挿入語句にどのような形容詞が付属するかについて表形式にして示した。また、どのような形容詞が挿入語句に付随するかという語の共起についてのデータを提示した。

6. 統計分析

6.1. 統計分析をおこなう理由

ここからは人手によって選定、分類した数値データを統計解析にかけて必要な結果を導き出す。一見複雑で意味の分からない数値処理を行なっているかのように見えるが、このような統計処理を行なうのには理由がある。

日々無数に発せられる言語情報についてある程度科学的に取り扱おうとする場合、他の多くの自然事象と同様に言語現象の多くは簡単に是非かに割り切れるものではない。100% そうだとは言いきれないが、何らかの傾向は存在することを示したいことがある。その場合、どの程度その傾向が確かかを論じたい場合に統計的な手法が用いられる。

統計学は主に二つに大分される：手元にあるデータについての情報を記述することを目的とする記述統計学（descriptive statistics）と、集めたデータから対象全体の状態を推測しようと試みる推測統計学（inferential statistics）である。

ここでは記述統計学だけではなく、推測統計学の手法も用いることにする。言語の特徴を明らかにする目的で研究活動を行なう場合、分析対象とする言語資料³⁸ は限られた量し

³⁸ 文学研究や言語研究の分野でそのような言語資料は「コーパス」と呼ばれる。「コーパス」という用語はコーパス言語学で用いられる品詞情報を付与した電子コーパスを特にさすことが多いが、言語研究や文学文化研究における言語資料全般のことを本来は意味する。また国立国語学研究所が公開している日本語のコーパスやロシアナショナルコーパスなどの言語データには品詞情報を表示するタグが付けられているが、本研究で使ったコーパスにはそのような処理は施されていない。

か手に入らない。手に入った言語データのみを問題とする記述統計学の手法では、手に入れた言語データの特徴しか分からない可能性がある。言語研究において「検定」「推定」などの手法は言語全体の姿を予測するために用いられる。

たとえば、ある程度の分量のコーパスを入手したとする。そこから何らかの傾向を見いだしたとしよう。ところが入手したコーパスは過去から現在に至るまで発話された言語のごく一部にすぎず、見いだした特徴も限られたコーパス内で偶然に見られるにすぎない可能性がある。そのような偶然の可能性に対処するために「検定」という作業が行なわれる。

本研究は言語としてのロシア語の全体像に対する関心から行なわれる。したがって、推測統計学の手法を用いる。

6.2. 統計分析とその結果

6.2.1. 用いる統計手法

以上の手続きで得られたデータを統計分析していく。考慮しなければならない問題は、数ある推測統計学の手法のうちどのようなものを用いるかである。本研究に適した手法を適用する必要がある。本研究の目的は、感情や評価を意味する挿入語句に主体や客体を表す語句が出現することと、作業仮説として挙げた次の3つの要因が関連しているかどうかを示すことである。その3つの要因は1) 挿入語句を構成する名詞の種類、2) 他の修飾語が付くか否か、3) 文体のちがひ、である。これらの3つの要因はそれぞれ性質を異にするため、それぞれ別の推測統計学の手法で分析する必要がある。

作業仮説1はどのような名詞が主体・客体の付与に関連するのかというものである。ここでは統計処理に特化したスクリプト言語Rを用いる。³⁹ 具体的に行なう作業としては、作業仮説を提起した際に列挙したものを人手によって頻度についての数値にし、下で詳述する χ^2 検定にかける。

本研究における作業仮説2と作業仮説3は、ある特徴がある要素の出現に関係・影響するのか、ということに帰着する。具体的には「ある特徴」とは（挿入語句の一部となっている名詞の種類、他の修飾語が付くかどうか、挿入語句のおかれる文体）のことであり、「ある要素」とは感情・評価を意味する挿入語句に付随して現れる мой や свой, его, ее などの修飾語のことを指す。関係・影響について取り扱う推測統計学における手法には「 χ^2 検定（独立性の）検定」と「相関分析」とが考えられる。相関分析は、2つの変数（表

本研究における「コーパス」とは単にネットから収集した電子化されたロシア語の文書群である。

³⁹ Rについては[<http://www.r-project.org>] (2014年9月1日閲覧), を参照。またRを用いた検定方法については、山田剛史・杉澤武俊・村井潤一郎『Rによるやさしい統計学』オーム社、2013年、128-137頁を参照した。

の横軸)の間の散らばりに正か負の相関関係を見いだそうとする。相関分析は2つ以上の個体(表の縦軸)がデータに存在する場合(サンプルサイズ)に使用されるのが普通である。また相関分析は最終的な研究結果を出す段階ではなく、予備的な分析として行なわれることが多い。⁴⁰ 今回は主体・客体を表現する要素が出現したか／しないかという2値的な性質のデータを取り扱う。そのため、本研究では独立性の検定(以降は先行研究にならって χ^2 検定と呼ぶ)を使用することとする。⁴¹

6.2.2. χ^2 検定とは

実際に収集し分析を施したデータが母集団(ロシア語全体)の特徴を反映しているか否か統計的検定によって確認する前に、今回使用する「 χ^2 検定」について説明する。⁴² χ^2 検定とは二つの変数の間に関連があるかどうか調べるための手法である。 χ^2 検定の有意判定は χ^2 分布という確率分布を用いておこなわれる。 χ^2 分布はデータ数によって異なった形をとる。つまり、ここでのデータ数は χ^2 分布の自由度を決定することになる。この χ^2 分布の自由度と棄却率によって有意か否かが判定される。この2つの変数は必ず二項であるように「肯定・否定」「有無」という形をもつことが多い。母集団(言語研究の場合、求めたい言語全体の状態)の分布が正規分布であると仮設した場合の仮説検定の手法の一つである。 χ^2 検定量 X^2 は次のように求められる:

$$\begin{aligned} X^2 &= (O_1 - E_1)^2 / E_1 + (O_2 - E_2)^2 / E_2 + (O_3 - E_3)^2 / E_3 + \dots + (O_k - E_k)^2 / E_k \\ &= \sum \{ (O_i - E_i)^2 / E_i \} \end{aligned}$$

ここで O_i は i 列目の実際に観測された値で実測度数と呼ばれる。 E_i は実測度数 O_i と対応して同じ列にあり、期待度数と呼ばれる。ここでデータは全部で k 列存在することを意味する。分子を二乗するのは負の値が出ないようにするためであると考えられ、検定に使用する χ^2 統計量は正より大きな値しかとらない。

χ^2 検定は、二つの変数の間に何の関連も存在しない(二つの変数は互いに独立である)という帰無仮説を統計的に棄却するために行なわれる。統計的検定は常に100%正しいものではなく、誤って有意であると結論される可能性が存在する。その誤っているかもしれ

⁴⁰ 石川慎一郎ほか『言語研究のための統計入門』くろしお出版, 2011年, 91頁。

⁴¹ ロシア語研究に χ^2 検定を使用した例として、秋山真一「主部に数詞句をもつ文の述語形態に関する歴史的研究」『ロシア語研究』第21号, 2008年, 19-49頁がある。本研究において χ^2 検定を実施する際に参考にした。

⁴² ここでの説明については、石川慎一郎ほか『言語研究のための統計入門』くろしお出版, 2011年, 60-68頁, を参考にした。

ないという可能性を表示するのが有意水準である。ここでは帰無仮説を棄却する有意水準は0.05（5%）とする。この有意水準とのちに詳しく述べる限定水準（以降はp値と呼ぶ）とをつきあわせて、帰無仮説を棄却するかどうかを決定する。

ピアソンの χ^2 検定は排他的な事象にしか用いることはできないことに注意する必要がある。すなわち1か0かの二進的な値に還元できるデータしか、検定にかけるとはできない、ということになる。

6.2.3. 挿入語句を構成する名詞の種類

どのような名詞が構成する挿入語句に主体・客体を表す語句が付随することが多いのかについて、粗頻度による情報を表形式にして提示する。

リソース 挿入語句	インターネット新聞	19世紀文学者の文書	集計
	主体・客体の明示／感情・評価を意味する挿入語句の数		
к счастью	0.00991 (8/807)	0.23456 (19/81)	0.03040 (27/888)
на счастье	0.50000 (2/4)	1.00000 (7/7)	0.81818 (9/11)
по счастью	0.00000 (0/17)	0.00000 (0/2)	0.00000 (0/19)
к несчастью	0.13793 (8/58)	0.19607 (10/51)	0.16513 (18/109)
на несчастье	0.00000 (0/0)	1.00000 (3/3)	1.00000 (3/3)
по несчастью	0.00000 (0/0)	0.00000 (0/0)	0.00000 (0/0)
к удовольствию	0.00000 (0/0)	1.00000 (6/6)	1.00000 (6/6)
к неудовольствию	0.00000 (0/1)	1.00000 (1/1)	0.50000 (1/2)
к изумлению	1.00000 (3/3)	0.75000 (36/48)	0.76470 (39/51)
к удивлению	0.90909 (30/33)	0.90151 (119/132)	0.90303 (149/165)
к радости	1.00000 (11/11)	1.00000 (18/18)	1.00000 (29/29)
на радость	1.00000 (7/7)	0.00000 (0/0)	1.00000 (7/7)
к сожалению	0.00967 (18/1861)	0.12000 (6/50)	0.01255 (24/1911)
к прискорбию	0.00000 (0/1)	0.50000 (1/2)	0.33333 (1/3)
к досаде	0.00000 (0/0)	0.66666 (10/15)	0.66666 (10/15)
к огорчению	1.00000 (1/1)	1.00000 (5/5)	1.00000 (6/6)
合計	0.03138 (88/2804)	0.57244 (241/421)	0.1020 (329/3225)

「幸運なことに」など肯定的な感情と「不運なことに」「残念なことに」など否定的な感情を意味する挿入語句において、主体・客体を表す語句が現れることは少ない。その一

方で、「驚いたことに」など驚きを意味する意外性の評価を意味する挿入語句において「誰が」を指し示す語句が頻繁に出てくるように感じられる。

6.2.3.1. クラスタリング

ここでの作業仮説は、挿入語句を構成する名詞によって、挿入語句に修飾語が付くか否かが決まるのではないかというものである。ここから、主体・客体がつくつかつかないかによって、上の表で挙げた挿入語句をグループ分けするという分析方法が考えられる。そこで、このグループ化のために、ここでクラスタリングを用いることにする。データのある特徴に着目して、似た特徴を持つものを一つのグループにまとめることによって分類を行う作業をクラスタリングと呼ぶ。クラスタリングは使用する手法によって結果が違ってくる。そのため、データの性質に合った手法を選択し、結果を人間によって解釈する作業が必要となる。

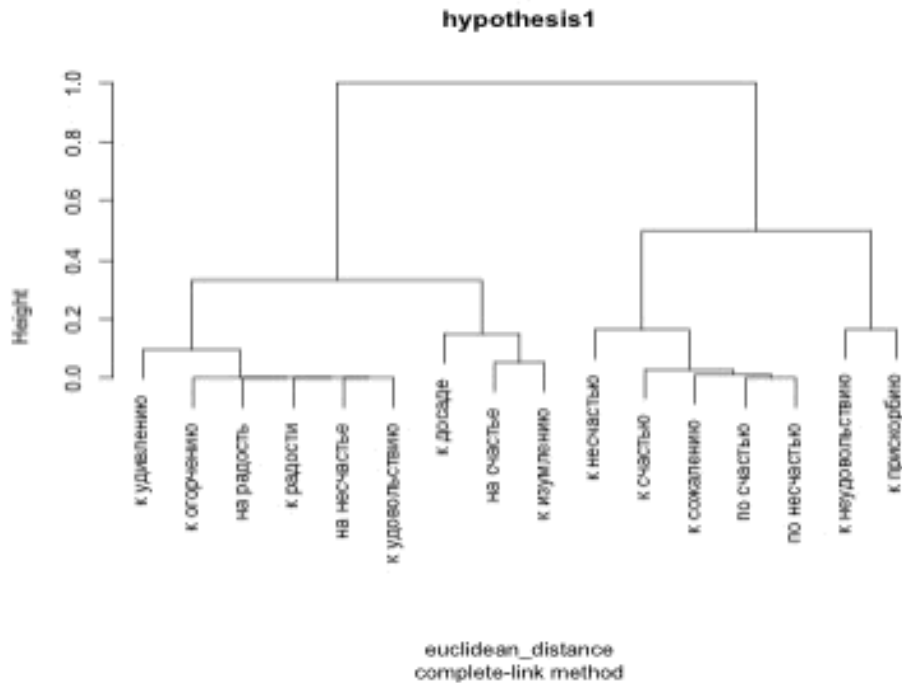
ここでの分析ではクラスタリング手法の中でも最も似ているもの同士をグループ化していく凝集型階層的クラスタリングを使用する。⁴³ クラスタの類似度は完全連結法 (complete-link method) によって計算した。完全連結法は最も基本的なクラスタリング手法の一つであり、比較する2つのクラスタのうち、最も距離の遠い要素の類似度を、それらのクラスタ間の類似度とする。完全連結法には、一個のクラスタが長くなりにくいという特徴がある。初期距離行列の作成にはユークリッド距離を用いる。クラスタリング分析にはRのメソッド `hclust()` を使用した。

6.2.3.2. 結果の解釈

インターネットのテキストにおける挿入語句に関してクラスタをおこなった結果は、次のようになる

⁴³ ここでのクラスタリングの説明については、高村大也『自然言語処理シリーズ1 言語処理のための機械学習入門』コロナ社、2010年、77-82頁、を参照した。

下で提示したクラスタリングの結果を解釈する。挿入語句 **на несчастье, по несчастью, к неудовольствию, к прискорбию** はサンプル数が少なかったため、解釈に使用する対象からは除外した。



挿入語句 **на радость, к радости, к удовольствию, к огорчению** は主体・客体が必ずついて現れることが明らかになった。その一方で、**по счастью** に主体・客体を表す語句が付く例は見られなかった。

挿入語句の中でも **к удивлению, на счастье, к изумлению, к досаде** は主体・客体とともに現れることが多い。それに対して **к счастью, к сожалению, к несчастью** といった挿入語句に主体・客体が付いて出現することは少ない。

ここでは名詞ごとに挿入語句と主体・客体の関係を分析した。その結果、どの挿入語句がよく用いられるかが明らかになった。具体的には、インターネットのニュース記事において **к сожалению** が、19 世紀文学において **к удивлению** が頻繁に見られる。これらの違いは報道の文体と文学作品の文体の違いに起因するものと考えられる。

6.2.4. 挿入語句に付随する修飾語

ここで取り扱う作業仮説は「他の修飾語の出現は主体・客体の出現に影響するかどうか」

ということである。ここで検討する挿入語句について、1) 主体・客体を表す語句と他の修飾語がともに現れる場合(例: к моему крайнему удивлению, к огромной радости местных жителей), 2) 主体・客体を表す語句のみが現れる場合(例: к радости ее, к удивлению моему, к изумлению публики), 3) 他の修飾語のみが付される場合(例: к величайшему прискорбию, к большому сожалению), 4) 主体・客体を表す語句も他の修飾語が付されることもない場合(例: к счастью, к сожалению), の4つの場合が考えられる。本研究ではこの点について着目し、人手で数え上げをおこなった。

この調査結果について χ^2 検定をおこなう。その場合、次のような仮説を設定した：

帰無仮説 H_0 ：「主体・客体の出現は他の修飾語の出現に関係がない」

対立仮説 H_1 ：「主体・客体の出現は他の修飾語の出現に関係がある」

これらの項目について χ^2 検定を実施できるようクロス集計表にまとめたものを以下に挙げる。すなわち 1) の場合は修飾語ありかつ主体・客体あり、2) の場合は修飾語なし主体・客体あり、3) の場合は修飾語あり主体・客体なし、4) は修飾語なしかつ主体・客体なしに対応する：

クロス集計表

	修飾語あり	修飾語なし
主体・客体あり	168	171
主体・客体なし	49	5712
合計	217	5883

このクロス集計表について補正なしのピアソン χ^2 検定にかけると、検定統計量は2213.865、自由度は1、p値は2.2e-16、という結果が得られた。p値である2.2e-16は2.2に0.1を16回かけた数値を意味し、有意水準0.05より小さいものとなり帰無仮説は棄却される。

ここでの調査と統計的仮説検定の結果、挿入語句における主体・客体の出現は、ほかの修飾語の出現と無関係ではないという結果を得ることができた。

6.2.5. ジャンルのちがい

この節で立証したい仮説は「文学作品かインターネット新聞かというジャンルのちがいによって、挿入語句の出現頻度も違ってくる」ということである。したがって、ここで棄却すべき帰無仮説は、「文学作品かインターネット新聞かというジャンルのちがいは挿入語句の出現に関連はない」となる。

帰無仮説 H_0 ：「感情・評価を意味する挿入語句に主体・客体を表す語句が付される頻度は、テキストのジャンルと関係がない」

対立仮説 H_1 : 「感情・評価を意味する挿入語句に主体・客体を表す語句が付される頻度は、テキストのジャンルと関係がある」

ここで「テキストのジャンル」とは本研究における言語分析に用いた 1) インターネット新聞と 2) 19 世紀文学者たちによる文書の 2 つを指す。

	インターネット新聞	19 世紀文学者の文書
主体・客体を表す語句あり	88	241
主体・客体を表す語句なし	2716	180
合計	2804	421

上の表に記載されている数値情報を χ^2 検定にかけ、次のような結果を得た。補正の実行は無効にしてある。検定統計量は 1259.19, 自由度は 1, p 値は 2.2e-16 となり, 有意水準 0.05 を下回るので帰無仮説は棄却される。

今回は 19 世紀文学者による文書とインターネット新聞という, ジャンルも時代区分も離れたものを比較した。19 世紀と 20 世紀, 現代という時代のちがいを考慮したものではない。また, ジャンルも新聞や文学作品だけではなく, 広告やソーシャルメディアの話し言葉もある。ジャンルのちがいだけではなく, 文学作品の中に存在する話し言葉の要素や文学作品特有の言い回しなどが検定結果に影響していることが考えられる。以上の理由から仮説の有効性は薄いと言わざるを得ないが, 本研究の取り扱う作業仮説の一つである「感情と評価を表す挿入語句に主体・客体を明示する語句がつくかつかないかには文体のジャンルも無関係ではない」ということを示唆することはできた。

6.2.6. 総括

この章では, 感情・評価を意味する挿入語句に主体・客体を表す語句がつくのどのような要因が関係しているのかという問題について, 1) 挿入語句を構成する名詞の種類, 2) 他の種類の修飾語がついているかどうか, 3) ジャンルのちがいという 3 つの仮説について調査した。その結果得られた頻度データを統計的仮説検定の一つ χ^2 検定にかけて, 集められたデータに見られる関連性がロシア語母集団においても有意であるかどうかを判定した。その結果, いずれの仮説についても有意であるという判定を得ることができた。ただし, 文体のジャンルに関する仮説についてはやや有効性が薄いと考えられる。

7. 結論と今後の課題

7.1. 結論

本研究では「感情・評価を意味する挿入語句に主体・客体を意味する語句がつくのは,

どのような要因が関係しているだろうか」という問題提起をした。そしてそれに対し 1) 挿入語句を構成する名詞の種類, 2) 他の種類の修飾語がついているかどうか, 3) ジャンルのちがい, という 3 つの要因を考えて検証を行なった。本研究の目的は, ロシア語学における貢献だけではなく, テキストマイニングを行なうときにも役に立つ情報を提供することであった。具体的な研究方法としては, 計算機によってネット上で得られたテキストデータから必要な文字列を検索して取得し, そこから人手で不適切なものを取り除き, 必要に応じて分類・整理を行なった。第一の仮説については, 挿入語句を構成する名詞の種類についてクラスタリングを行なった。残りの 2 つの作業仮説を数値データの形に整理して, 関連性がロシア語全体において存在するかどうかを調べるための統計的仮説検定の手法の一つである独立性の検定 (χ^2 検定) にかけて, 有意であるかどうかを調査した。

第一の作業仮説を検証するために, 挿入語句のクラスタリングをおこなった。その結果, 挿入語句 на радость, к радости, к удовольствию, к огорчению は主体・客体が必ずついて現れることが明らかになった。その一方で, по счастью については主体・客体なしのサンプルしか見られなかった。主体・客体が付いた例と付かない例の両方の場合が見られた挿入語句については к удивлению, на счастье, к изумлению, к досаде は主体・客体とともに現れることが多い。それに対して к счастью, к сожалению, к несчастью といった挿入語句に主体・客体が付いて出現することは少ない。

統計的仮説検定を使用する作業仮説 2 と 3 については, すべてにおいて有意であるという結果が得られた。ただし, 文体に関する第三の仮説については課題が残された。

本研究では, 感情・評価を表す挿入語句の表す意味が「誰のもので」「誰・何に対して」発せられたのかを表す修飾語を「主体・客体」と名付け, それらの出現に影響する要因について調査・考察をおこなった。1) 挿入語句 на радость, к радости, к удовольствию, к огорчению は主体・客体が必ずついて現れることが明らかになった。その一方で, по счастью に主体・客体を表す語句が付く例は見られなかった。挿入語句の中でも к удивлению, на счастье, к изумлению, к досаде は主体・客体とともに現れることが多い。それに対して к счастью, к сожалению, к несчастью といった挿入語句に主体・客体が付いて出現することは少ないことが明らかになった。また, 挿入語句に主体・客体が付くか付かないかに影響する他の要因として, 2) ほかの修飾語の有無, 3) 文体のジャンルの違い, が関係していることが明らかになった。

7.2. 残された問題と今後の課題

7.2.1. 残された問題

本研究で解決することのできなかった問題を挙げる。今回取り扱った作業仮説のうち、ジャンルの問題についてはもう少し詳細に検討する必要があると思われる。人手による言語分析を行ったときの印象であるが、感情・評価を意味する挿入語句のあるものは話し言葉において出現することが多い。話し言葉において、文脈で明らかな語はしばしば省略されるという要因が考えられる。本研究においてこの問題を取り扱うことはできなかった。この話し言葉と書き言葉による影響を調べることは、今後の課題となる。

また今回コーパスを 19 紀文学作品だけではなく、4 つのインターネット新聞に広げて文例を集めた。およそ 22 万の文書を検索したことになる。集めてみて明らかになったことだが、分析対象とした感情・評価を意味する挿入語句というものは、それほど頻繁に用いられない。したがって、この挿入語句をそのまま素性として実際のテキストマイニングに応用することは難しいと言わざるを得ない。

本研究では、作業仮説 2 と 3 の検証に際して、基礎的な統計的検定法である χ^2 検定という手法を用いた。この検定方法は、個々の作業仮説が母集団について妥当であると推測されるかどうかのみを意味しているに過ぎない。例えば、2～3 の仮説で示された要因のうち、どれが強く作用しているかを統計的に明らかにする性質のものではない。この問題について調べるには、別の適切な統計手法を使用する必要がある。

7.2.2. 今後の課題・方針

最後に、現段階で考えられる本研究の今後の方針を列挙していく。第一に考えられる今後の方針として、テキストマイニングへと応用を可能とする成果を出していくことが考えられる。本研究はロシア語をよりよく理解する助けとなる情報を提供することを目的として行なわれたが、その先の応用事例の一つとして、ソーシャルメディア上のロシア語を分析することが考えられる。例えばロシア語で書かれたマイクロブログのテキストをマイニングすることによって、日露関係に有益な情報を発見する、などが考えられる。

このためにも、今回より大きなサイズのテキストデータを分析・分類する必要がある。今回は、人手によって検索された言語データを分析した。しかし、今後はすべてのデータを人手で分析・分類することは現実的に不可能である。そのため、現段階では言語データを自動的に分類する分類器を構築することを考えている。このことを実現するために、大量のデータを処理できる統計学の手法、とりわけ教師あり学習の手法をロシア語のデータに応用することが、第二の課題として挙げられる。

付録 Web スクレイピングに使用した python コード(<http://www.pravda.ru/>から記事を自動的に取得するためのものです。python はブロックをインデントによって表現することに注意してください。環境によってエスケープ文字を適宜書き換えてください)

```
# -*- coding:utf-8 -*-      # utf-8 方式エンコーディングであることを指示

import re,nltk,urlib2, socket,os  # 必要なライブラリをインポートする

# url から html データを取得する
def page_open(url):
    o = urllib2.urlopen(url)
    html = o.read()
    o.close()
    return html

# 正規表現によって必要な情報を抽出する
def extract_data(regex,html):
    l = []
    for i in re.findall(regex,html):
        l.append(i)
    return l

# 処理したテキストデータをファイル形式で保存する
def _write(txt,date,i):
    w = open(date+i+".txt","w")
    w.write(" "+txt)
    w.close()

lenta_head = "http://www.pravda.ru/news/" # ヘッダ情報を抽出する正規表現
regex_eachpage = ¥
re.compile('<a class="articleAnons" href="(http://www.pravda.ru/news/[a-z0-9¥W]+)/">¥n')
# 記事の url を取得する
def get_lenta(num):
    lenta_url = lenta_head+str(num)+"/"
```

```

lenta_html = ""      # html データを入れておく変数を用意
lenta_html = page_open(lenta_url) # 記事の場所を示す url の html データを取得
lenta_page = []      # 個々の url を入れておくリストを初期化
lenta_page = extract_data(regex_eachpage, lenta_html) # url をリストにして取得
return lenta_page    # url を入れたリストを返す

regex_date = re.compile('{\d{2}-\d{2}-\d{4}}') # 日付を取り出すための正規表現
# 日付情報を正規表現によって抽出する
def get_date(lenta):
    retdate = ""      # 数であっても正規表現で抜き出されるのは文字列型となる
    for date in re.findall(regex_date, lenta):
        retdate = date[-4:] + date[3:5] + date[0:2] # 数値のみを文字列として格納
    return retdate

# html 形式のデータからテキストデータを抜き出すための正規表現
first_paragraph = re.compile('<!-- первый абзац -->([А-Яа-я0-9ЎWЇwЇs]+)<div
class="social-text-in-material" style="margin:0;padding:0 0 15px 0;"><p><strong>Читайте
также:</strong></p>')
text_align = re.compile('^\n<p style="text-align: justify;">([А-Яа-я0-9ЎWЇwЇs]+)Читайте
также:')
# 不要部分を取り除くための正規表現
erase_txt = re.compile('Читайте также:([А-Яа-я0-9ЎWЇwЇs]+)')

# html 形式のデータからテキストデータを抜き出す
def get_article(sent_html):
    html, txt = "", ""
    html = page_open(sent_html)
    for h2 in re.findall(first_paragraph, html):
        txt += h2
    for align in re.findall(text_align, html):
        txt += align
    txt = nltk.clean_html(txt)
    return re.sub(erase_txt, "", txt) # 不要部分を取り除いたものを文字列として返す

```

```

# エントリーポイント
if __name__ == "__main__":
    begin = 0    # 取得する記事の url の開始番号
    end = 1000   # 取得する記事の url の終了番号
    import time
    lenta_page = []    # 記事の置いてある url を格納するリスト
    article = ""       # 記事を格納する文字列変数
    num = 100         # ファイル番号を初期化する
    tmpdate = "0"      # 日付が変わっていないか確認するための変数
    for i in range(begin, end):
        try:
            # 記事の置き場所をリスト形式で取得する
            lenta_page = get_lenta(i)
            # url を開いて記事と日付を受けとる
            for txt_url in lenta_page:
                senddate = get_date(txt_url)
                if senddate != tmpdate:    # 日付が変わっていないかチェック
                    num = 100              # 日付が変わるとファイル番号を降り直す
                    tmpdate = senddate
                article = get_article(txt_url) # html ファイルから記事を取り出す
                _write(article,senddate,str(num)) # 記事をテキスト形式で保存する
                num += 1                    # ファイル番号をインクリメント
                time.sleep(3)               # サイトに負荷をかけないためのディレイ
        # エラーと例外を捕捉する
        except urllib2.URLError, e:
            pass
        except socket.error, e:
            pass
        except socket.timeout, e:
            pass
        except httpplib.IncompleteRead, e:
            pass

```

Имплицитный субъект и объект вводных элементов с эмоциональным и оценочным значением

СЭРИ Акинори

Цель настоящей статьи — выяснить правила проявления показателя субъекта и объекта в вводных словах и предложениях русского языка. В статье используются методы индуктивной статистики и техники компьютерной лингвистики.

Показатели субъекта и объекта вводного слова на практике иногда существуют, а иногда отсутствуют: *На счастье его, в воротах опять прошло благополучно*; *Кисида, по его словам, не считает, что Япония подпадает под это эмбарго*. В вводном предложении *на счастье его* местоимение *его* указывает, кто счастлив, а в предложении *по его словам* — кто сказал. Этот показатель в тексте иногда сокращен: *К счастью, она, моя голубушка, не узнала его*. Тут возникает следующий вопрос: какие правила существуют в проявлении показателя субъекта и объекта вводных слов и предложений? Данная работа посвящена объяснению правил проявления слов, которые указывают на субъект и объект вводного слова и предложения.

К разрешению этого вопроса могут быть привлечены следующие факторы: лексические, синтаксические и стилистические. С помощью статистических методов выбран преимущественней из перечисленных выше факторов.

Данное исследование характеризуется двумя следующими особенностями: 1) Для анализа используется корпус (данные для языкознания), собранный из текстов интернета. Источники данных включают в себя не только тексты художественной литературы, но и тексты электронных медиа; 2) Для поиска языковой информации в тексте и ее статистического анализа применяются специально написанные компьютерные программы. Используется также сценарный язык программирования с библиотекой индуктивных статистических подпрограмм.