

## ロシア語テキストへのメタ情報タグ付与に関する一考察

阿出川 修嘉

### An Approach to the Determination of Metadata Tags for Russian Texts

In this article, we examined the meta information which is added to texts in the Russian National Corpus (RNC) and examined what kinds of metadata can be added in order to collect required samples more efficiently.

In the current RNC, the morphologic and semantic metadata is generally attached to a unit of meaning classified as “word”. By using such metadata, users can collect various samples by setting composite search criteria. However, meta information is not attached to the unit of meaning classified as “sentence”. Therefore, there is room for improvement in order to obtain more meaningful search results. In this paper, we reviewed the possibility of adding additional metadata to the text constituting the RNC corpus by making use of such linguistic features of Russian language as sentence classification.

Russian sentences can be classified into 5 types according to their morphological and semantic features. These are: 1) the definite personal sentence (*Я люблю работать*), 2) the indefinite-personal sentence (*Здесь говорят только по-русски*), 3) the generalized personal sentence (*Тише едешь, дальше будешь*), 4) the impersonal sentence (*Мне холодно!*) and 5) the infinitive sentence (*Тебе этого не понять!*). If metadata classified in this way is attached to the “sentence” unit, it will be possible to investigate the grammatical behavior of words in more detailed way.

The difficulty in attaching such metadata is that sentence classification cannot be performed mechanically for all sentences. Therefore, when classifying, it is necessary to consider their semantic context. However, for some sentence types, it is possible to determine the classification unambiguously. Firstly, if the sentence is accompanied by a grammatical subject (i.e. a nominative noun), the sentence can be classified as a ‘definite personal sentence’. On the other hand, if the sentence is not accompanied by a grammatical subject, it is necessary to differentiate between cases of non-past tense and past tense. In the case of

non-past tense, the classification of sentences can be unambiguously defined when the verb is in the form of the first-person singular. In this case, it can be marked mechanically as a 'definite personal sentence'. In the case of the past tense, it is only when the form of the verb is a male form or a female form, that it can be defined as a 'definite personal sentence'. Whenever a sentence has an impersonal predicate, it is unambiguously determined as an 'impersonal sentence'.

## 0. はじめに

言語研究において、言語形式が見せる文法的な振る舞いの観察・分析にコーパスが利用されるようになったのは、とりわけ英語学の分野を中心に既に久しいが、現代ロシア語<sup>1</sup>を対象とする研究においてもそうした傾向は同様である。本国ロシアにおいて、ロシア語を対象とする大規模なコーパス構築の進展が2000年代に入って本格化したこともあり、コーパスを利用したロシア語の分析は近年一層活発化している<sup>2</sup>。

ロシア語を対象とするコーパスとして、現時点で規模的に最も大きく、また現在も拡張が続けられているのは、『ロシア語ナショナルコーパス (Национальный корпус русского языка; Russian National Corpus)』である<sup>3</sup>。様々なジャンルのテキスト<sup>4</sup>が網羅されており、その総語数は現時点で3.6億語を超えている<sup>5</sup>。

このコーパスは、それ以前のコーパスで可能だった、文字列による単純な検索に加え、語の形態論情報などを利用した検索（後述）も可能となっており、言語研究者が必要とするサンプルを効率よく収集できるようになっている。

しかしながら、ロシア語という言語の特徴を十分に踏まえた形での検索が自在に可能というわけでは必ずしもなく（後述）、その点では、総語数の拡充という以外にも、まだ発展の余地が残されていると考えられる。

本稿では、このロシア語ナショナルコーパスを対象として取り上げつつ、このコーパスではどのような調査が可能で、どのような点について不足があるか、またその不足している点を踏まえた上で、コーパスを構成しているテキストに今後どのようなメタ情報を付加していけば、どのような調査が可能になるかという可能性について考察する。

## 1. ロシア語ナショナルコーパスの文法情報タグについて

### 1.1 概要

本節では、「ロシア語ナショナルコーパス」(以下 RNC と略記する) を構成するテキストに付与されている、各種メタ情報のタグについて確認し、これらのメタ情報を用いて抽出可能なサンプルについて確認することしよう。

上でも述べたように、RNC は、ロシア語を対象としたコーパスとしては現時点で規模の観点からも、また内容面から見ても、最も充実したコーパスであり、現在もなお拡張が試みられているモニターコーパスである。

このコーパスを構成している言語資料(テキスト)には、以下のようなメタ情報が付加されている。

- ① 形態論に関わる情報<sup>6</sup>
- ② 語彙的意味に関わる情報<sup>7</sup>

これらの情報(特に形態論について)を付与する作業は、先行して存在していた種々のプログラムを用いて実現されている<sup>8</sup>。

以下で、上記のそれぞれについて見ていくことにしよう<sup>9</sup>。

### 1.2 形態論的情報

RNC で付与されている形態論的情報は、大きく分けて、1) 品詞情報に関わるものと、2) 文法カテゴリーの情報に関わるものに分けることができる。

下表が品詞情報に関わる情報の一覧である。

表 1：RNC のメタ情報タグ：品詞情報に関わるもの

記号 (英)	記号 (露) <sup>10</sup>	品詞	具体例 <sup>11</sup>
S	сущ	名詞 (существительное)	яблоня, корпус
A	прил	形容詞 (прилагательное)	коричневый, морской
NUM	числ	数詞 (числительное)	четыре, десять; много
ANUM	числ-прил	形容詞的数詞 (числительное-прилагательное)	один, седьмой
V	глагол	動詞 (глагол)	читать, пользоваться
ADV	н	副詞 (наречие)	очень
PRAEDIC	предик	述語 (предикатив)	хорошо, жаль, пора
PARENTH	вводн	挿入語 (вводное слово)	кстати, по-моему
SPRO	мс-сущ	代名詞 (местоимение-существительное)	она, что
APRO	мс-прил	代名詞的形容詞 (местоимение-прилагательное)	твой; который
ADVPRO	мс-н	代名詞的副詞 (местоименное наречие)	где, вот
PRAEDICPRO	мс-предик	述語代名詞 (местоимение-предикатив)	некого, нечего
PR	предлог	前置詞 (предлог)	под, напротив
CONJ	союз	接続詞 (союз)	и, чтобы
PART	част	小詞 (частица)	бы, же, пусть
INTJ	межд	間投詞 (междометие)	увы, батюки

次に、文法カテゴリーに関する情報の一覧を確認しておこう。

表 2：RNC のメタ情報タグ：文法カテゴリーに関わるもの

文法カテゴリー	記号 (英)	記号 (露)	カテゴリー名
性 (род)	m	м	男性 (мужской род)
	f	ж	女性 (женский род)
	m-f	м-ж	共通性 (общий род)
	n	с	中性 (средний род)
有生性 (одушевленность)	anim	од	活動体 (одушевленное)
	inan	неод	不活動体 (неодушевленное)
数 (число)	sg	ед	単数 (единственное число)
	pl	мн	複数 (множественное число)

文法カテゴリー	記号 (英)	記号 (露)	カテゴリー名
格 (падеж)	nom	им	主格 (именительный)
	gen	род	属格 (родительный)
	dat	дат	与格 (дательный)
	dat2	дат2	与格 2 (дательный) <sup>12</sup>
	acc	вин	对格 (винительный)
	ins	твор	具格 (творительный)
	loc	предл	前置詞格 (предложный)
	gen2	род2	属格 2 (родительный) <sup>13</sup>
	acc2	вин2	对格 2 (винительный) <sup>14</sup>
	loc2	предл2	前置詞格 2 (предложный) <sup>15</sup>
	voc	зв	呼称形 (звательный)
adnum	сч	基数形 (счётная форма) <sup>16</sup>	
形容詞の 短語尾・長語尾	brev	кратк	短語尾形 (краткая форма)
	plen	полн	長語尾形 (полная форма)
比較級・最上級 (степень сравнения)	supr	превосх	最上級 (превосходная)
	comp	сравн	比較級 (сравнительная)
	comp2	сравн2	比較級 2 (сравнительная) <sup>17</sup>
体 (вид)	ipf	нсв	不完了体 (несовершенный)
	pf	св	完了体 (совершенный)
他動性 (переходность)	tran	перех	他動詞 (переходный глагол)
	intr	нп	自動詞 (непереходный глагол)
態 (залог)	act	действ	能動態 (действительный залог)
	pass	страд	受動態 (страдательный залог)
動詞の諸形 (репрезентация глагола)	inf	инф	不定形 (инфинитив)
	partcp	прич	形動詞 (причастие)
	ger	деесп	副動詞 (дееспричастие)
法 (наклонение)	indic	изъяв	直說法 (изъявительное наклонение)
	imper	повел	命令法 (повелительное наклонение)
	imper2	повел2	命令法 2 (повелительное наклонение) <sup>18</sup>
時制 (время)	prate	прош	過去 (прошедшее)
	praes	наст	現在 (настоящее)
	fut	буд	未来 (будущее время)
人称 (лицо)	1p	1л	1人称 (1-е лицо)
	2p	2л	2人称 (2-е лицо)
	3p	3л	3人称 (3-е лицо)
その他	persn	имя	固有名詞 (имя собственное)
	patrn	отч	父称 (отчество)
	famn	фам	名字 (фамилия)
	zoon	зоон	動物の呼び名 (кликка животного)
	0	0	不変化詞 (несклоняемое)

### 1.3 語彙的意味に関わる情報

次に、意味論に関する情報について確認しよう。RNCでは、以下の5品詞の語に対して、語彙的意味に関わる情報が付されている。

- ① 名詞
  - (ア) 物を表す名詞 (предметные имена)
  - (イ) 物以外を表す名詞 (непредметные имена)
  - (ウ) 固有名詞 (имена собственные)
- ② 形容詞
- ③ 数詞
- ④ 動詞
- ⑤ 副詞

上記のうち、名詞に関しては、「物を表す名詞」、「物以外を表す名詞」及び「固有名詞」の三つに下位区分が成されている。

また、それぞれ、語彙的意味に関する注記 (лексико-семантические пометы) として、「分類 (таксономия)」、「部分 (мереология)」、「評定 (оценка)」に関する注記が付されている。また、「語形成 (словообразование)」に関する注記が付されている場合もある。

下表では、この語彙的意味に関する情報のうち、「物を表す名詞」に分類されるものの一部分を提示してある。

表 3：語彙的意味に関わる情報：分類（一部）

記号 (英)	品詞	具体例
t:hum	人 (лица)	человек, учитель
t:hum:etn	民族名 (этнимы)	эфиоп, итальянка
t:hum:kin	親族名称 (имена родства)	брат, бабушка
t:hum:supernat	超自然物質 (сверхъестественные существа)	русалка, инопланетянин
t:animal	動物 (животные)	корова, жираф, соколка, ящерица, муравей
t:plant	植物 (растения)	береза, роза, трава
t:stuff	物質 (вещества и материалы)	вода, песок, тесто, желье, шелк

記号 (英)	品詞	具体例
t:space	空間・場所 (пространство и место)	космос, город, тайга, овраг, вход
t:constr	建造物 (здания и сооружения)	дом, шалаш, мост
t:tool	道具 (инструменты и приспособления)	молоток, палка, пуговица, машина

下表では、語彙の意味に関わる情報のうち、「部分」に関するものの一部を示してある。

表 4：語彙の意味に関わる情報：部分（一部）

記号 (英)	品詞	具体例
pt:part	部分 (части)	верхушка, кончик, половина
pt:partb & pc:hum	身体の部分, 人体器官 (части тела и органы человека)	голова, сердце, ноготь
pt:partb & pc:animal	動物の身体・器官の部分 (части тела и органы животных)	хвост, жало
pt:part & pc:plant	植物の部分 (части растений)	лист, ветка, корень
pt:part & pc:constr	建造物の部分 (части зданий и сооружений)	комната, дверь, арка
pt:part & pc:tool	設備の部分 (части приспособлений)	деталь, лопасть, крышка
pt:part & pc:tool:instr	道具の部分 (части инструментов)	топорище, лезвие
pt:part & pc:tool:device	機械の部分 (части механизмов и приборов)	дисплей, корпус, кнопка
pt:part & pc:tool:transp	輸送手段の部分 (части транспортных средств)	руль, колесо, капот
pt:part & pc:tool:weapon	武器の部分 (части оружия)	дуло, курок, эфес

「評定 (оценка)」に関する注記とは、ある語が肯定的な評定の意味を担っているか (例えば умница のような語)、否定的な評定の意味を持っているか (例えば негодяй といった語)、あるいはそうした評定の意味を含んでいないかを表す。

「語形成 (словообразование)」に関する注記とは、ロシア語の語形成規

則によって形成された語かどうかを示す。例えば, писатель や создатель といった語には, それぞれ動詞 писать, создать から派生した語であるということを示すタグが付される。

#### 1.4 RNC におけるメタ情報タグの具体例

前節までで, 形態論に関わる情報, 意味論 (語彙的意味) に関わる情報の双方について, どのようなメタ情報が付されているのかについて確認した。ここでは, 実際の例を確認しておこう。

RNC では, 検索によって得られたテキストを構成している語のひとつひとつに対して, 以下のような情報が付されており, 任意の語をクリックすることで, 検索結果の画面上でそれらの情報を語ごとに確認することができる。

- ① レンマ (лемма)
- ② 文法的情報 (грамматика)
- ③ 主要な意味 (семантика основная)
- ④ 補足的意味 (семантика дополнительная)
- ⑤ その他の補足の特徴 (дополнительные признаки)

①の「レンマ」は, 当該の語の辞書の見出し形を示し, ②の「文法的情報」では, その語形が当該テキスト内でどのような文法的意味(cf. 第 1.2. 節)を示しているか, ③の「主要な意味」では, 語彙的意味に関わる情報 (cf. 第 1.3. 節) が付されている。④の「補足的意味」では, 語形成に関わる情報などが示される。最後の⑤の「その他の補足の特徴」で示されているのは, その後が文頭に位置している, あるいは文の最後に位置しているなどの情報が示される。

例として以下の文を取り上げてみる。

- (1) Снега ещё нет, но запах в воздухе уже витает.

雪はまだないが, 大気中に香りはもう漂っている。

(Андрей Геласимов. Фокс Малдер похож на свинью. 2001)

この文を RNC の検索にかけると, 以下のようなメタ情報が付された形

で出力される。

表 5：RNC で得られるメタ情報

語	付されているメタ情報				
	レンマ	文法的情報 <sup>19</sup>	主要な意味	補足的意味	補足的特徴
Снега	снег	сущ, неод, м, ед, род, disamb	r:abstr, r:concr, t:stuff, t:weather	-	capital, first
ещё	ещё	н, disamb	-	-	-
нет	нет	предик, disamb	-	-	bcomma, bmark
но	но	союз, disamb	-	-	acomma, amark
запах	запах	сущ, неод, м, ед, им, disamb	der:v, hi:class, pc:tool:cloth, r:abstr, r:concr, t:smell, t:space	-	-
в	в	предлог, disamb	-	-	-
воздухе	воздух	сущ, неод, м, ед, предл, disamb	r:concr, t:stuff	r:concr, t:space, t:stuff	-
уже	уже	н, disamb	t:time	-	-
витает	витать	глагол, нп, нсв, изъяв, действ, наст, 3л, ед, disamb	ca:noncaus, d:root, t:move	-	bdot, bmark, last

このように、一語に対して、レンマ、形態論的情報などのメタ情報が付されている<sup>20</sup>。

### 1.5 RNC で現状可能な調査 (RNC の基本検索)

前節までで、RNC において付与されているメタ情報タグについて確認した。

これらのメタ情報が付されているコーパスでは、冒頭にも述べたとおり、様々な検索条件の設定が可能であり、研究者が求める言語資料を得ることができる。

現時点では、基本検索においては以下のような検索が可能である。

- ① 語形による検索 (поиск точных форм)
- ② 語彙・文法検索 (лексико-грамматический поиск)
  - (ア) レンマ

- (イ) 文法カテゴリーの情報 (грамматические признаки)
- (ウ) 語彙的意味に関する情報 (семантические признаки)
- (エ) 補足的な特徴 (дополнительные признаки)
- (オ) 語形成に関する情報 (словообразование)

上記①の語形による検索では、単純に任意の語形（あるいは語句）を検索語（検索キー）として検索を行うと、その語形（語句）がどれくらい用いられているかについての統計情報と、それが用いられているテキストと共に結果が表示される。

表 6：RNC での検索例（1）：語形による検索

検索キー	検索結果
книгу	語形 книгу を含む文

上記②の語彙・文法検索では、あるレンマ（上記②－ア）に対して、文法カテゴリーの情報（上記②－イ）、語彙的意味に関する情報（上記②－ウ）、補足的な特徴（上記②－エ）、語形成に関する情報（上記②－オ）をそれぞれ組み合わせた形で検索を行うことができる。下表の例では、слово という語（レンマ）が複数・属格（形態は слов となる）で用いられている例を調べるといふ検索キーと、それによる検索結果を示している。

表 7：RNC での検索例（2）：レンマ，文法カテゴリー

検索キー		検索結果
レンマ ②－ア	文法カテゴリーの情報 ②－イ	
слово	属格 (gen), 複数 (pl)	слов を含む文

また、レンマを指定せずに、その他のメタ情報に関する諸条件を組み合わせることで検索は可能である。例えば、下表で示されている例は、生き物の意味を表す女性名詞が、複数形且つ対格形で用いられている物の中から（＝形態論に関わる条件）、肯定的な意味を有する物（＝意味論に関わる条件）を調べるといふ目的を満たすものである。

表 8：RNC での検索例（3）：メタ情報のみの組み合わせ

検索キー			検索結果
レンマ (②-ア)	文法カテゴリーの情報 (②-イ)	語彙の意味に関する情報 (②-ウ)	
-	名詞 (S), 対格 (acc), 複数 (pl), 女性 (f), 活動体 (anim)	物を表す (r:concr), 肯定的 (ev:posit)	героинь, дочек など

しかしながら、ロシア語の言語的特徴を利用した検索方法というものを考えてみると更に他の方法での検索を模索することができそうである。次節ではその点について考えてみよう。

## 2. ロシア語文の分類に関わる情報

### 2.1 概要

本節では、RNC においてまだ実現できていない検索方法のオプションについて考えてみよう。

前節で見たように、現時点での RNC は、「語」という単位に対する形態論的あるいは意味論的な情報についてはかなりの程度整備されてきており、それに基づいて様々な検索が可能になっていることが分かる。

しかしながら、その「語」が、どのような言語的環境で現れてくるのかという点について調査をしようと思うと困難が生じる。例えば、ある任意の動詞（例えば работать）の文法的行動を調べたいと思った場合に、どのような形態で用いられているかに関するサンプルは検索・収集できる（работать, работает, работал, работаю などの諸形態）が、その動詞がどのような言語的環境、例えばどのような種類の文中で用いられているかということまでは検索によっては調べることができない。ここで念頭に置いているのは、ロシア語の文の統語論上の分類と、文で用いられている動詞（あるいは名詞でも）の使用との関係である。後述するように、ロシア語の文はその統語論的・意味論的特徴に応じて分類が可能であり、全ての文は原則としてその分類に当てはめることができる。したがって、この分類についてのメタ情報がコーパスに加えられていれば、テキストにおける語の文法的な振る舞いがより詳しく調査できると考えられるのである。

本節では、まずロシア語文の分類について概観し、またそれに伴ってロシア語動詞のパラダイムについても確認する。次に、こうした文の分類に関するメタ情報を付加した場合の利点について考察し、実際にこれらのメ



場合には、主語の「数（単数及び複数）」と「人称（1～3人称）」の別に応じて、合計6通りの形態がある<sup>25</sup>。下表は、動詞 *читать* の非過去時制の諸形態である。

表 9：ロシア語動詞のパラダイム（非過去時制）

人称	単数	複数
1人称	читаю	читаем
2人称	читаешь	читаете
3人称	читает	читают

それに対して、過去時制の場合には、主語の「数（単数及び複数）」と「性（男性・女性・中性）」の別に応じて異なる。なお、「性」の別に応じて形態の対立があるのは、主語が単数の場合のみで、主語が複数の場合には「性」の別に応じた形態の対立は中和してしまい、形態的には一つの形態になる（下表）。

表 10：ロシア語動詞のパラダイム（過去時制）

単数		複数
男性	читал	читали
女性	читала	
中性	читало	

これらを踏まえた上で、次節でロシア語文のそれぞれの分類の特徴について確認していこう。

### 2.2.2 それぞれの文の統語論的及び意味論的特徴

まず、「定人称文」は、主語（主格の名辞類<sup>26</sup>）と動詞述語を有する文、もしくは動詞述語のみで構成される文である。

- (4) Иван                                    читает                                    книгу.  
 John-M-SG-NOM                    read-IMPV-PRS-3SG                    book-F-SG-ACC  
 イワンは本を読んでいる。





### 2.3 文の分類タグを付加すると可能になること

前節で、ロシア語の文の分類について確認した。これらの分類を、コーパスのサンプルのメタ情報タグとして付加することが可能になると、どのような利便性が得られるだろうか。ここではその点について考えてみよう。

上で見たとおり、現時点で可能な検索は、語形による検索及び語という単位ごとに、形態論的情報と意味論的情報とを組み合わせる形での検索である。

RNCから取り出した以下の例文を例に見てみよう。いずれの例文も、ウェブ上のフォーラムでのやり取りから抜き出されたものである。

- (14) Так что, все же собираетесь провести следующие тысячу двести лет в попытках **понять** свое собственное существование?

だから、自分自身の存在を理解しようとして、次の1200年を同じように過ごすつもりなのか？

[Forum: 12 часов в день? Не могу согласиться с М. Прохоровым (2010-2011)]

- (15) Просто это нужно **понять** и принять свою точку зрения.

ただそのことを理解して自分の見解を受け入れる必要がある。

[Forum: 12 часов в день? Не могу согласиться с М. Прохоровым (2010-2011)]

- (16) Сколько с вами общаюсь, не могу **понять**, то ли вы хорошо притворяетесь, то ли знатный демагог...

どれだけあなたたちとやり取りしても、あなたが上手いこと取り繕っているのか、それともデマゴグなのかを理解できない。

[Forum: 12 часов в день? Не могу согласиться с М. Прохоровым (2010-2011)]

- (17) Да и мне, пожалуй, никогда не **понять**, чем заводятся сердца чувственных велосипедистов, оттого и не дерзаю сравнивать.

おまけに、私は恐らく他国の自転車競技者の心臓が何によって動かされているのかかつて一度も理解できたことがないので、比較するこ

とができない。

[Форум: Горный двухподвесочный (2010)]

これらは全て同じ **понять** という動詞不定形（上記例文中でボールドにしてある）を含む文である。しかし、不定形の用いられている使用環境は異なっている。例（14）は、**собираетесь**（2人称複数）を述語とする定人称文で、不定形は名詞 **попытка** と語結合を成している。例（15）は、無人称述語 **нужно** を述語とする無人称文であり、不定形は無人称述語の補語的な役割を果たしている。例（16）は動詞 **мочь**（例文中では1人称単数の形態を取っている）を述語とする定人称文である。不定形は述語動詞と語結合を成し、補語的な役割を果たしている。例（17）は、**понять** が単独で述語として機能しており、不定法文に分類される。

これらの使用環境の違いは、下表のようにまとめられる。

表 12：понять の使用環境の違い

例	文の分類	不定形が用いられている統語環境
(14)	定人称文	名詞 ( <b>попытка</b> ) との語結合
(15)	無人称文	無人称述語 ( <b>нужно</b> ) との語結合
(16)	定人称文	動詞 ( <b>мочь</b> ) との語結合
(17)	不定法文	述語

現時点で可能な **понять** という語形による検索だけでは、これら四つの例は数多くのサンプルの中に混在することになってしまうが、テキストに文の分類に関するメタ情報が付されていれば、これら四つの例は、文の分類に関する検索条件をそれぞれ設定することによって、異なるサンプルとして取り出すことが可能になると考えられるのである。

## 2.4 分類タグの付加の方法と作業上の障壁

### 2.4.1 タグを付加する単位

上で見たように、ロシア語という言語の特徴を加味した上で、文の分類に関する情報を表すタグがテキストに付加されていれば、コーパスから様々なサンプルを得ることができそうである。しかしながら、現時点まで、このメタ情報タグがテキストに付与されていないのは、何らかの障壁

があるために実現できていないのだろうと予想される。本節では、具体的にどのように分類タグを付与していけるか、またその際にどのような障壁に遭遇すると考えられるかについて考察してみよう。

ここで、問題となるメタ情報タグのセットについて確認しておこう。

- ① 形態論的信息（「語」という単位に対する）
- ② 語彙的意味に関わる情報（「語」という単位に対する）
- ③ 「文の分類」に関わる情報（「文」という単位に対する）

従来のメタタグ（上記①及び②）は、いずれも「語」という単位に対して既に付与されているものである（cf. 第1.2.節）。それに対して、ここで新たに提案している「文の分類」に関するメタ情報タグ（同③）は、「文」という単位に付与されるという点で、その性質を大きく異にしていることが分かる。

#### 2.4.2 タグをどのように付与していくか

問題となる「文の分類」に関わるタグを付加する際に求められる作業としては、大まか以下のような作業が考えられるだろう。

この「文の分類」に関わるタグは、「文」という単位に付加するので、まずテキストにおける「文」という単位の確定をする必要がある。したがって、まず、文の開始となる語と終了となる語をマークアップするという作業が必要となってくる。これについては、上でも見たとおり、RNCにおいては「補足的な特徴」の項目で、文頭に来る語には、「capital」や「first」というマークアップがなされており、また文末に来る語には「last」のようにマークアップされている。このため、文という単位の確定には、この「first」というタグから「last」というタグまで、というように既存のマークアップをある程度まで利用することは可能だろう<sup>28</sup>。

「文」という単位が確定できた後に、その「文」がどのタイプに分類されるかについての情報を、その「文」を構成する語に付与していけば、その文字列が「文」であり、且つあるタイプの「文の分類」に属しているということが分かる。

例えば下の例文を取り上げて、具体的に考えてみよう。

(18) Я люблю работать в школе.

私は学校で働くのが好きなのです。

[Форум: Мужчина в школе (Взгляд на Мужчину в школе снаружи и изнутри) (2011)]

上で (表5を参照) も確認したように, 現時点ではレンマ, 文法的情報, 主要な意味, 補足的意味, 補足の特徴という, 五つのメタ情報が「語」という単位に付与されている。下表は, これらのメタ情報に, 本節で検討している「文の特徴」という項目を加えたものである。

表 13 : 文の分類についてのメタ情報を付与した例

語	レンマ	文法的情報	主要な意味	補足的意味	補足の特徴	文の特徴
Я	я	мс-сущ, 1л, ед, им, disamb	r:pers	—	capital, first, numred	定人称文 (文頭)
люблю	любить	глагол, перех, нсв, изъяв, действ, наст, 1л, ед, disamb	ca:noncaus, d:root, t:psych:emot	d:root	numred, persred	定人称文
работать	работать	глагол, нп, нсв, инф, действ, disamb	der:s	der:s	posred	定人称文
в	в	предлог, disamb	-	—	-	定人称文
школе	школа	сущ, неод, ж, ед, предл, disamb	r:concr, t:org	r:abstr, t:fam	bdot, bmark, last	定人称文 (文末)

文の特徴として, 文の分類を, その文で用いられている語の全てに付するのは煩雑ではあるが, 検索に際して何らかの利用上の便宜が期待できるので, 付しておくのが適切であるように思われる。

#### 2.4.3 タグを付与する際の障壁

前節では, 主語を含んでいる「定人称文」を取り上げ, 新たに文の分類についてのメタ情報を付与する作業について考えてみた。

しかしながら, 文の分類に際して想像される障壁は, その「文」が, どのタイプに分類されるかについての確定を巡る問題である。端的に言えば, 実際のテキスト内の文の分類の全てを, 機械的に, 一義的に定めることが

必ずしもできないという点にまとめられるだろう。大量のテキストデータを処理するということを考慮すれば、機械的に行える部分はそうしてしまう方が作業の効率は高まっていくが、実際のテキストにおいては、そのように単純に処理できる例ばかりではない。

上で見た例(18)では、文の分類自体は、1人称の主語を含む文であるから定人称文であると確定できるが、問題となるのは、主語を含まない文の場合である。注20で見た、下の文を改めて例にとって確認してみよう。

(19) Ночевали на чердаке пустой заброшенной сторожки на слежавшихся, потерявших запах листьях.

積もり積もって香りを失った落ち葉の上に建った、人も住んでおらず荒れるに任せた番小屋の屋根裏部屋で夜を明かした。

[Юрий Казаков. «На охоте» 1956.]

この文は、主語がなく、動詞の形態(Ночевали)を見ると、過去時制で複数形となっている。この文の分類の可能性としては、「定人称文」と「不定人称文」の双方の場合がありうる。そして、そのどちらになるかの判断は、(広義の)文脈を確認した上で最終的に可能になる。

この文に関して言えば、後の文脈で登場人物達(父と子)の描写が続くので、ここで動詞が複数形となっているのは、主語としてこの主人公達(=3人称複数)が念頭に置かれているからであると判断できる。したがって、この文の分類は、主語の省略された「定人称文」ということになる。

このように、本来ある主語が省略されたものなのか、それとも普遍人称文、不定人称文のように、そもそも主語が想定されていないタイプの文なのかという判断は、形式的な基準のみによって機械的に行うには困難を伴うものになる。

## 2.5 部分的な解決に向けて

### 2.5.1 分類が一義的に定められるものを検討する

上で、ロシア語文の分類の可能性とそれに伴う困難について見た。

しかしながら、一義的に文の分類を定められるケースが部分的にでも存在するのであれば、それらのマークアップ作業から始めることができるだろう。ここでは、このような、文の分類を行っていく上での現時点で可能

な作業について、次善の策を考えてみることにする。

まず、主語（主格の名辞類）を含む文については、これらは文法的な主語があることで定義上人称が定まるので、これは一義的に定人称文とマークアップすることができる<sup>29</sup>。

それに対して、主語を含まない文の場合、上でも見たとおり、五つの異なる分類に振り分けられる可能性を考慮しなければならない。これら主語を含まない文の分類を考える際に手がかりとなるのは、動詞の形態と文全体の意味である。

効率よく（機械的に）文の分類を確定していくためには、必ずしも言語形式との一対一の対応関係が保証されていない意味に依るのではなく、何らかの言語形式に基づいて一義的に定められるというのが最も理想的な確定の方法である。

動詞の形態は、先に見たとおり（cf. 第 2.2.1. 節）、非過去時制の場合と過去時制の場合とでパラダイムが異なるため、当該文の時制に応じて考慮する必要があるということになる。非過去時制の場合には、数と人称のカテゴリーに応じて形態が変化するので、合計 6 形態がある。それに対して、過去時制の場合には、数と性のカテゴリーに応じて変化するので、合計 4 通りの形態がある。これらは下表のようにまとめられる。

表 14：ロシア語動詞の諸形態

	文法カテゴリー	形態
非過去時制	数(単数・複数) 人称(1～3人称)	читаю, читаешь, читает; читаем, читаете, читают
過去時制	数(単数・複数) 性(男性・女性・中性)	читал, читала, читало; читали

以下で、主語を含まない文の場合についてそれぞれ検討してみよう。

## 2.5.2 主語を含まない文：定人称文の場合

まず、主語を含まない定人称文のケースについて考えてみよう。

まず非過去時制の場合について考えてみよう。上でも見たとおり、非過去時制の場合には、動詞は主語の数と人称に応じて変化する。動詞が 1 人称（単数・複数）の形態の場合には、定人称文である。2 人称（単数・複数）の場合は、定人称文の場合と普遍人称文の場合とがあり得るため、一義的

には定められない。同様に3人称単数の場合も、定人称文の場合と無人称文の場合とがあり得る。複数の場合には、定人称文の場合と不定人称文の場合とがあり得るため、やはり一義的には定められない。

これらをまとめると下表のようになる<sup>30</sup>。

表 15：動詞の形態と文の分類【非過去時制】

動詞の形態		文の分類
数	人称	
単数	1人称	定人称文
	2人称	定人称文, 普遍人称文
	3人称	定人称文, 無人称文
複数	1人称	定人称文
	2人称	定人称文
	3人称	定人称文, 不定人称文

この場合、一義的に定められるのは、動詞の形態が1人称単数の場合であるということになる。

次に、過去時制の場合について考えてみよう。動詞は、主語の数と性のカテゴリーに応じて変化する。動詞が男性形及び女性形であれば、1人称の代名詞（я）及び2人称の代名詞（ты），及び3人称代名詞のうち、男性もしくは女性の代名詞（он, она）が省略されていると考えられるので、それは定人称文であると考えてよい。動詞が中性形の場合には、中性名詞が主語となっている場合（その場合には「定人称文」と）、無人称文の場合との双方があるので、一義的には定められない。複数形の場合には、名詞の複数形が主語となっている場合と、不定人称文の場合とがありうるため、やはりこれも一義的に定めることは困難である。

これらをまとめると下表のようになる。

表 16：動詞の形態と文の分類【過去時制】

動詞の形態		文の分類
数	性	
単数	男性	定人称文
	女性	定人称文
	中性	定人称文, 無人称文
複数		定人称文, 不定人称文

したがって、過去時制の場合、一義的に文の分類を定められるのは、動詞の形態が男性形及び女性形の場合ということになる。

### 2.5.3 主語を含まない文：不定人称文の場合

不定人称文は、上でも少し触れたように、非過去時制の場合であれば、3人称複数の動詞形態、過去時制であれば複数形の動詞形態によって表現される（下例；再掲）。

(6) В Сибири не любят горячку и спешку.  
 in Siberia-F-SG-LOC not love-IMPF-PRS-3PL rush-F-SG-ACC and haste-F-SG-ACC  
 シベリアでは、せっかちは好まれない。

(7) В поселке построили новую школу.  
 in settlement-M-SG-LOC build-PF-PST-PL new-SG-F-ACC school-F-SG-ACC  
 村に新しい学校が建てられた。

上の例(19)を見ても分かるとおり、これらの動詞形態は、定人称文の場合にも現れうるため、動詞の形態から一義的に不定人称文と確定するのは困難である。

### 2.5.4 主語を含まない文：無人称文の場合

無人称文の場合には、述語の別に応じて考慮する必要があるだろう。

述語が無人称述語に分類されるものであれば、文の種類は一義的に無人称文と分類することが可能である。したがって、そのような述語を含む文は直ちに無人称文とマーキングして良いということになる。RNCに関し

例えば、無人称述語には、「PRAEDIC (предик)」という文法情報に関するマーキングが既に成されている (cf. 第 1.2. 節) ので、文中にこれが含まれるのであれば、その文は無人称文とマークアップすることが可能である。

(20) Завтра            будет            холодно.  
tomorrow        be-PRS-3SG    it\_be\_cold  
明日は寒くなるでしょう。

(21) Вчера            было            холодно.  
yesterday        be-PST-SG-N    it\_be\_cold  
昨日は寒かった。

述語が動詞の場合には、いくつかの解釈の可能性が生じてくる。無人称文の場合、動詞の形態は、非過去時制の場合には 3 人称単数、過去時制であれば中性形が選択される。下例は、無人称動詞 знобить (悪寒を感じさせる) が用いられている例である。

(22) Его            знобит.  
he-ACC        make\_smb\_feel\_feverish-IMPF-PRS-3SG  
彼は寒気を感じている。

(23) Его            знобило.  
he-ACC        make\_smb\_feel\_feverish-IMPF-PST-SG-N  
彼は寒気を感じていた。

しかしながら、上表を見て分かるとおおり、この二つの形態 (非過去時制での 3 人称単数、過去時制での中性形) が用いられるのは、無人称文の他にも、定人称文の場合 (下例) もありうるため、やはり一義的には定められない。

- (24) Письмо лежало на столе.  
 letter-N-SG-NOM lie-IMPF-PST-3SG-N on table-M-SG-LOC  
 その手紙はデスクの上にあった。

### 2.5.5 主語を含まない文：不定法文の場合

次に不定法文について考えてみよう。

不定法文とは、述語として不定形が用いられている文だが、そのことを機械的に確定するのは困難である。第一の条件として、「文の中に不定形が含まれている」という条件がひとつ考えられるが、容易に想像が付くとおり、これはその文が不定法文であることの必要十分条件とはならない。というのも、ひとつの文の中において、不定形が動詞述語の補語的な役割で用いることができ、その場合には、定人称文と分類される（下例）。

- (25) Люблю читать.  
 love-IMPF-PRS-1SG read-IMPF-INF  
 読書が好きです。

また、同一文中に動詞の定形が含まれているサンプルを排除できたとしても、以下のような文が得られ（ここでは不定形が名詞と語結合を成している）、これもやはり文の分類としては人称文に分類される。

Угрозы Запада отказать от российских  
 threat-F-PL-NOM Eastern refuse-IMPF-INF from Russian-ADJ-PL-GEN  
 Europe-F-SG-GEN

энергоносителей — это попытка нас  
 energy this is attempt-F-SG-NOM we-ACC  
 carrier-PL-GEN

запугать, оказать на нас давление.  
 frighten-PF-INF give-PF-INF on we-ACC pressure-N-SG-ACC

(Вадим Пономарев, Дмитрий Кобылкин. Сила России // «Эксперт», 2014)

西欧側のロシアのエネルギー事業者を拒絶するという脅しは、我々を驚かせ、我々に圧力を与えようという試みである。

これらは、いわゆる不定形の統語論的用法のうち、「接語的用法」として位置付けられているものである (cf. Грамматика 2005: § 2745)<sup>31</sup>。

このように、ある文を不定法文であると確定するためには、最低限の条件として、1) 文の中に不定形が含まれており、且つ、2) 同一文の中で、他に動詞の定形が含まれていない、という二つの条件を満たすものをまずサンプルとして抽出し、その上で更に、3) そこで用いられている不定形が接語的用法ではない、という条件を用いてフィルターにかける必要があるだろう。

しかしながら、それを可能にするためには、ある文中で用いられている不定形が、接語的用法なのか非接語的用法なのかについてのメタ情報タグを付するという、別の作業が必要となってくる。

### 3. 本稿における結論に代えて：現実的な解

本稿では、RNCにおけるメタ情報タグの内容と、現状可能な検索などについて確認し、より効率よく必要なサンプルを収集するために、更にもどのようなメタ情報を付加することができるかを考察した。

現時点のRNCでは、「語」という単位に対して付されている形態論的、意味論的なメタ情報はかなりの程度充実した物となっており、それらを利用して複雑な条件を設定してコーパスを検索、サンプルを収集することが可能となっている。

しかしながら、その一方で、「文」という単位に対してはメタ情報が付けられておらず、その点では未だ改善の余地があることが分かった。そこで本稿では、ロシア語の言語的特徴を活かしてより有益な検索結果を得るための試みとして、伝統的なロシア語の文の分類を利用する方法について考察した。すなわち、定人称文、不定人称文、普遍人称文、無人称文、不定法文といった分類を、コーパスを構成しているテキストに付与するという方法である。このような、「文」という単位に対するメタ情報の付与が実現すれば、ある語が、どのような文において用いられるかという、より詳細な語の文法的振る舞いについての調査が可能になる。

しかし一方で、そうしたメタ情報の付与には相応の困難が伴うことも明

らかになった。最大の問題点としては、全ての文に対して、文の分類の情報を、形態論的な指標のみに基づいて機械的に付与することができないという点があげられる。

ただ、いくつかのケースに関しては、一義的に文の分類を確定することは可能である。こうしたケースに当てはまるデータに対してメタ情報を付与していく作業にまず着手するというのが現実的なアプローチとなるだろう。

一義的に分類を確定できるケースとして、まず、文が文法的な主語を伴っている場合が考えられ、その場合には当該文は定人称文と分類できる。

それに対して、文が文法的な主語を伴っていない場合には、非過去時制と過去時制の場合のそれぞれについて考慮する必要がある。非過去時制の場合、一義的に文の分類を定められるのは、動詞が1人称単数の形式の場合である。この場合には、機械的に定人称文としてマークアップすることができる。

主語の有無及び動詞の形態（非過去時制の場合）の差異に応じた文の分類をまとめると、下表のようになる。

表 17：動詞の形態と文の分類（非過去時制）

動詞の形態		主語のある文	主語のない文
数	人称		
単数	1人称	定人称文	定人称文
	2人称	定人称文	定人称文, 普遍人称文
	3人称	定人称文	定人称文, 無人称文
複数	1人称	定人称文	定人称文
	2人称	定人称文	定人称文
	3人称	定人称文	定人称文, 不定人称文

それに対して、過去時制の場合には、動詞の形態が男性形と女性形の場合に限っては、定人称文と認めることができる。同様に、下表のようにまとめることができる。

表 18：動詞の形態と文の分類（過去時制）

動詞の形態		主語のある文	主語のない文
数	性		
単数	男性	定人称文	定人称文
	女性	定人称文	定人称文
	中性	定人称文	無人称文
複数		定人称文	定人称文、不定人称文

なお、文の述語として（動詞ではなく）無人称述語が用いられている場合には、無人称文と定められるので、その文の分類も一義的に定めることが可能である。

ここまで見てきたように、全ての文に対する、言語形式のみによる機械的なメタ情報付与は困難ではあるが、一部の文については一義的に文の分類を付与することが可能であることが分かる。

#### 4. 今後の課題

ここまで、RNCのコーパスに付与されているメタ情報について確認し、更に高度な検索を可能にするために、どのようなメタ情報を加えることができるかについて考察してきた。しかしながら、本稿における考察及び提案はまだ着想の段階にあるため、今後各作業段階の細部について検討を重ねていく必要はあるだろう。

まず、テキスト処理の具体的な手順についての考察が現状では不十分であることは否めない。そのため、大量のテキストの生データを取り扱う際に、技術的に可能かどうか、また効率的なサンプル収集という目的に照らした場合に、今回の作業手順がどの程度有効に機能するかどうかについては別途実際に検証を行わなければならない。

また、本稿では試みに、伝統的なロシア語学における「文の分類」を取り上げ、それをメタ情報として利用する方法について考察を加えたが、この「文の分類」の規定の難しさも、今後分析を進めていくことで解消を図っていかなくてはならない点だろう。上でも述べたように、形態論的、統語論的な特徴のみからでは文の分類は確定しきれず、意味的な側面も多分に関わってくる。このような、意味的な要素と形式的な要素との相関関係について、より詳細に分析を行っていく必要がある。

また、「語」という単位に付しうる他のメタ情報についても、その付与可能性について今後検討していく価値はあるだろう。例えば、本文でも触れたような、動詞の不定形の「接語的用法」、「非接語的用法」といった分類はそのひとつと考えられるだろう。これらの用法の差異に応じてメタ情報を付することができるのであれば、サンプル収集の際に一層利便性を高めることができると思われる。これらのメタ情報も利用できれば、ある動詞が、どのような統語環境で、どれくらい用いられているのかなどについてより詳細な調査が可能になる。

今後は、主にこれらの課題を克服するべく、更なる分析及び作業を進めていく必要があるだろう。

## 5. 文献

- Грамматика 2005 — Русская грамматика. Репринтное издание. Гл. ред. Н.Ю. Шведова. М., 2005.
- «Как пользоваться Корпусом?» (URI: <http://www.ruscorpora.ru/instruction-main.pdf>)  
[記載の URI は 2017 年 9 月時点で有効]
- Результаты и перспективы 2005 — Национальный корпус русского языка: 2003-2005. Результаты и перспективы. М., 2005.
- Русский язык-Теория 2006 — Современный русский язык. Теория. Анализ языковых единиц. Под редакцией Е.И. Дибровой. М., 2006.
- Труды 2004 — Труды международной конференции «Корпусная лингвистика — 2004»: 11-14 октября 2004 г. СПб., 2004.
- 雲越繁 (2005) : 『ロシア語ナショナルコーパス (Национальный корпус русского языка) について』, ロシア語研究会年報『ロシア語研究』, 第 18 号, pp. 49-70。

## 注

- 1 以下単に「ロシア語」とする。
- 2 そのような研究の成果がまとめられたものとして、Результаты и перспективы (2005) や Труды (2004) などがある。
- 3 トップページ URI は以下：

<http://www.ruscorpora.ru/>

このコーパスに関して、技術的側面に重きが置かれている初期の解説としては、雲越 (2005) があるのでそちらも参照されたい。

- 4 以下の URI の情報による。コーパスを構成するテキストについての詳細も以下に詳述されている：

<http://www.ruscorpora.ru/corpora-structure.html>

- 5 コーパスを構成している、それぞれのジャンルのテキストに関する統計なども以下

の URI に記述されている：

<http://www.ruscorpora.ru/corpora-stat.html>

- 6 詳細は以下の URI で確認することができる：

<http://www.ruscorpora.ru/corpora-morph.html>

- 7 詳細は以下の URI で確認することができる：

<http://www.ruscorpora.ru/corpora-sem.html>

- 8 以下の URI でそれらのプログラムについて（特に MyStem と Dialing）述べられている。

<http://www.ruscorpora.ru/corpora-progr.html>

- 9 RNC では、本文で取り上げた二つのメタ情報（形態論、意味論）に加え、「統語論に関する情報」についてのタグも付加されているが、ここでは、基本検索で利用することのできる「文法的情報」と「意味論的情報」の二つのメタ情報に限り以下論を進める。「統語論的情報」のタグについての詳細については下の URL を参照されたい：

<http://www.ruscorpora.ru/instruction-syntax.html>

- 10 ロシア語版の記号（略号）については、筆者が見たところウェブページ内には明確に記載されておらず、筆者が検索結果から適宜調べたものである。
- 11 ここで示す具体例は、当該ページに掲載されているものに基づいているが、複数挙げられているもののうち、スペースの都合によりいくつか割愛したものがある。また、筆者の判断でより平易な語を加えている場合がある。
- 12 「по несколько раз（何回かずつ）」のような語結合で見られる、分配の意味を表す与格を指している。
- 13 「чашка чаю（カップ 1 杯分の紅茶）」のような語結合で見られる чаю のような、いわゆる数量を表す属格のことを指している。
- 14 「по два человека」のような語結合で見られる два が念頭に置かれている。
- 15 「в лесу」のような語結合で用いられる лесу のような前置詞格を指す。
- 16 「два часа」における「часа」のように、数詞との結合に際して用いられる形態が念頭に置かれている。
- 17 「поглубже」のような、接辞 по と形容詞・副詞の比較級との結合によってできた語を指している。
- 18 「Идёмте!」の例に見られるような、ある種の動詞の 1 人称複数形態に、接辞 те を付して形成される動詞の形態が念頭に置かれている。
- 19 ここで提示される文法的情報の略語は、コーパスのインターフェイスの言語に応じてロシア語と英語との二通りの提示が可能となっている。すなわち、英語のインターフェイスを用いていれば、これらの情報は英語の略語によって提示される。
- 20 ただ、全てのテキストに対して、このような適切に整備されたメタ情報が付与されているわけではない。形態論的に複数の文法的意味を示しうる、同綴異義（омонимия）のケースが解決されていないテキストが存在している。例えば、下例を取り上げてみよう。

Ночевали на чердаке пустой заброшенной сторожки на слежавшихся, потерявших

запах листьев.

積もり積もって香りを失った落ち葉の上に建った、人の住んでいない、荒れるに任せた番小屋の屋根裏部屋で(彼らは)夜を明かした。

(Юрий Казаков. «На охоте» 1956.)

このテキストについては、同綴異義が解決されておらず、文中の語 запах に対しては、男性名詞 запах の対格とみなすメタ情報 (сущ, неод, м, ед, вин, disamb) と、同じく名詞の主格とみなすメタ情報 (сущ, неод, м, ед, им, disamb)、また動詞 запахнуть の過去形 (単数, 男性形) とみなすメタ情報 (глагол, нп, св, изъяв, действ, прош, м, ед) が付与されており、可能性のある全ての情報が付加されて表示される (実際には、このテキスト内の запах は、名詞の対格である)。このように、現時点では、全てのテキストについて、メタデータの整備の工程が完了しているわけではないようである。なお、このような同綴異義語の複数の解釈の可能性が排除されているものについては、「омонимия снята」(英語のインターフェイスの場合には「disambiguated」) と付記されており、そうでないものについては「омонимия не снята」(同「not disambiguated」) と付記されている。

- 21 これは、伝統的なロシア語統語論では、それぞれ前者が「二肢文 (двусоставное предложение)」、後者が「一肢文 (односоставное предложение)」とそれぞれ呼ばれているものとはほぼ等しい。
- 22 以降例文にはグロスを付するが、グロスは原則として Leipzig Glossing Rules に習っている。また議論に照らして不要と思われる箇所には付していない場合もある。本稿で用いている略号は以下の通り：

1:1 人称、2:2 人称、3:3 人称、ACC:対格、ADJ:形容詞、ADV:副詞、DAT:与格、F:女性、FUT:未来形、GEN:属格(生格)、IMP:命令形、INS:具格(造格)、IPFV:不完了相(不完了体)、LOC:所格(前置格、前置詞格)、M:男性、N:中性、NOM:主格、PFV:完了相(完了体)、PL:複数、PRS:現在形、PST:過去形、SG:単数

なお、Leipzig Glossing Rules については、以下の URI を参照されたい：

<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>

- 23 厳密には、ここで挙げた物の他にも、『名辞文 (номинативное предложение)』(「Весна!」(春が来た!))のような名詞だけで構成されている文)と『呼びかけ文 (вокативное предложение)』(「Маша!」(マーシャ!))のような人名だけで構成されている文)という種類の文がある。本稿では、動詞述語と文の分類との関わりについての考察を優先するため、ここでは差し当たり議論の俎上には載せない。
- 24 「不定形」は、辞書の見出し語としても用いられており、ある動詞を示す際の基準となる形態として用いられる。また、ロシア語学では伝統的に「不定形」という訳語が当てられているが、これは「定形」に対する術語であり、その意味で採用する理由は相応にあると考えられ、以下本稿ではこの伝統的な訳語を採用している。
- 25 なお、ロシア語の代名詞はこの三つの人称と二つの数に応じて用意されており、以下のような体系を成している。

人称	性	単数	複数
1 人称	—	я	мы
2 人称	—	ты	вы
3 人称	男性	он	они
	女性	она	
	中性	оно	

3 人称単数の代名詞は、性に応じて三種類の代名詞が用意されているが、同じ 3 人称でも複数になると性による形態的な対立はなくなる。

- 26 ロシア語の統語論上の分類では、「主語 (подлежащее)」と分類されるのは、名詞・代名詞類の主格のみである。
- 27 「普遍人称文」は、本文中で述べたように動詞の形態は 2 人称単数のものが典型的である。しかしながら、その統語構造は様々で、定人称文の場合 (「Семеро одного не ждут.」七人は一人を待たない。) や、不定法文の場合 (「Волков бояться—в лес не ходить.」オオカミが怖けりゃ森へは入るな。), 命令形の場合 (「Семь раз отмерь, один — отрежь.」七度測って一度裁て。) などもあり、この普遍人称文というタイプの文は、形態・統語論的特徴のみでは確定しきれないのが実情であり、今後さらなる分析を進めていく必要がある。本稿では差し当たり、典型的なケースである、動詞が 2 人称単数の物を対象と絞り、以下論を進める。
- 28 ただ、この既存のマークアップが利用できるのは、当該文字列が単文の場合のみであり、接続詞などが用いられて複文となっている場合には利用できないため、そうしたケースの場合には、当該文字列が単文あるいは複文の一部であるというマークアップを別途用意するなどの対応を考える必要がある。
- 29 ただ、下例のようなタイプの文は、例外的なケースとして扱わなければならない。

Меня	зовут	Саша.
I-ACC	call-IMPF-PRS-3PL	Sasha-M-SG-NOM

私はサーシャと言います。

この例では、固有名詞の主格が現れているが、これは文法上の主語ではなく動詞の補語として機能しており、この文の分類自体は不定人称文となる。

- 30 Русский язык-Теория (2006) でも、同様に動詞の形態と文の分類について考察されており (cf. Русский язык-Теория 2006: 289), ここでの議論においても参考にしている。
- 31 「接語的用法 (присловное употребление)」とは、他の語と (統語的位置に依ることのない不変的な) 語結合を成して用いられる用法である。Грамматика(2005) では、1) 情報補完機能 (「Обязан **знать** 知っている義務がある」, 「мастер **рассказывать** 物を語るのが得意な人」のような例), 2) 客体の意味 (「научить **читать** 読み方を教える」のような例), 3) 定語の意味 (「выйти **погулять** 散歩に出る」のような例) に下位区分されている。それに対して、「非接語的用法 (неприсловное употребление)」

とは、統語的位置に大きく影響を受ける語結合に基づく用法であり、1) 述語の特徴の意味（「Она хохотать. 彼女は急に笑い出した。」のような例）、2) 主体的意味（「Сломать дерево—не шалость. 木を砕くのは容易ではない。」のような例）、3) 客体的意味（「Стало быть, это ты ему приносила пить и есть, Нелли? 大方、彼に飲み物と食べ物を選んでやったのは君だろう? ネリー。」のような例）、4) 定語の意味（「Принес книгу почитать. 読む本を持ってきた。」のような例）に下位区分が成されている。なお、ここでの語の結合における統語的位置による影響などについては、Грамматика (2005) 参照されたい (cf. 2005: § 1720)。また、「接語的用法」及び「非接語的用法」という訳語は、筆者による試訳である。