

*Академия управления
при Президенте Республики
Беларусь*

Шибут М.С., Яковичин В.С.

**РАСПОЗНАВАНИЕ
ГРАММАТИЧЕСКОЙ ИНФОРМАЦИИ
В ПРОЦЕССЕ ФОРМИРОВАНИЯ
ЛИНГВИСТИЧЕСКИХ ЗНАНИЙ**

Задача оперативного формирования специализированных баз знаний

Специализированные базы знаний:

- ограниченные некоторой предметной областью
- полученные на основе обработки определенного вида электронных ресурсов – материалов форумов, блогов ...

Особенность – индивидуализированные языки участников диалогов:

- использование ненормативной и эрративной ненормативной лексики (языков в стиле эрратива)

В таких условиях использование статических лингвистических знаний (универсального грамматического словаря) затрудняется

1. Словарь окончаний - 1

Окончания	Типы словоизменения
а	$f_1-f_4, m_1-m_7, n_1-n_4$
ам	$f_1-f_5, m_1-m_7, n_1-n_4$
ами	$f_1-f_5, m_1-m_7, n_1-n_4$
ат	c_7
ах	$f_1-f_5, m_1-m_7, n_1-n_4$
ая	a_1-a_6
е	$f_1-f_4, f_8, f_9, m_1-m_{11}, n_1-n_3, n_5-n_7$
его	a_3, a_6, a_7
ее	a_3, a_6, a_7
ей	a_3, a_6, a_7
ем	$a_3, a_6, a_7, c_1-c_6, m_5-m_{12}, n_3-n_8$
ему	a_3, a_6, a_7
ет	c_1-c_6
...

- $f_1-f_7 (f_1, \dots, f_7)$ – склонение существительных женского рода;
- $m_1-m_{12} (m_1, \dots, m_{12})$ - склонение существительных мужского рода;
- $n_1-n_8 (n_1, \dots, n_8)$ - склонение существительных среднего рода;
- $a_1-a_7 (a_1, \dots, a_7)$ – адъективное склонение;
- $c_1-c_{12} (c_1, \dots, c_{12})$ – спряжение

2. Алгоритм выявления коррелятивных пар

1. $i := 1$;
2. $j := i + 1$;
3. если $a_i = s - P$, шаг 4; иначе шаг 5;
4. если $s = I(a_i)$, шаг 8; иначе - шаг 11;
5. выделение предполагаемой основы $I(a_i, a_j)$ и окончаний $F(a_i)$ и $F(a_j)$;
6. если $|I(a_i, a_j)| \geq 2$; $|F(a_i)|, |F(a_j)| \leq 3$, шаг 7; иначе - шаг 11;
7. если для $F(a_i)$ получено множество парадигм $P(a_i)$, шаг 8; иначе - шаг 11;
8. если для $F(a_j)$ получено множество парадигм $P(a_j)$, шаг 9; иначе - шаг 11;
9. если $P_i \cap P_j \neq \emptyset$, шаг 10; иначе шаг 11;
10. $a_i := s - P$; удалить элемент a_j ;
11. $j := j + 1$;
12. если $j \leq n$, шаг 3; иначе шаг 13;
13. $i := i + 1$;
14. если $j \leq n$, шаг 2; иначе КОНЕЦ работы алгоритма.

a_1, \dots, a_n - массив словоформ

$I(a_i, a_j)$ – предполагаемая общая основа;

$F(a_i)$ и $F(a_j)$ – предполагаемые окончания словоформ

$P(a_i)$ и $P(a_j)$ – множества парадигм словоизменения

$s - P$ – общая основа s , снабженная множеством общих типов парадигм P

3. Преобразование

$a_i = \text{стола},$

$a_j = \text{столов},$

$P(a_i) = \{f_1-f_4, m_1-m_7, n_1-n_4\}$

$P(a_j) = \{m_1-m_3\};$

$P(a_i) \cap P(a_j) = \{m_1, m_3, m_3\}$

<i>стола</i>		<i>стол-т2</i>
<i>столбец</i>		<i>столбец</i>
<i>столетий</i>		<i>столетии-п7, п8</i>
<i>столетию</i>	\Rightarrow	\emptyset
<i>столов</i>		\emptyset
<i>столового</i>		<i>столов-а2</i>
<i>столовый</i>		\emptyset
<i>столы</i>		\emptyset

4. Выявление согласования

$a(f_1-f_4) \rightarrow 1.113$

окончание *-а*, используемое в типах женского склонения f_1-f_4 , заменяется значениями «существительное», «именительный падеж», «единственное число», «женский род» (*карта, книга*);

$a(m_1, m_6) \rightarrow 1.211/120$

окончание *-а*, используемое в типах мужского склонения m_1, m_6 , заменяется значениями «родительный падеж», «единственное число», «мужской род» (*дома, мастера, сторожа*) или значениями «именительный падеж», «множественное число» (*домá, мастерá, сторожá*)

Правила согласования:

3.120.X₁ и 1.120.X₂

большие дома, новые книги

3.51.X₁ и 1.511.X₂

большим домом, новым столом

5. Алгоритм распознавания значений управления

1. морфологический анализ (кодирование значений согласования);
2. выявление и свертывание синтагм, основанных на согласовании;
3. выделение в оставшихся словосочетаниях гипотетических синтагм;
4. верификация гипотетических синтагм и пополнение списка новыми элементами;
5. обработка списка синтагм (кодирование значений управления).

5. Алгоритм распознавания значений управления

Писатель подарил ученикам новую книгу с дарственной надписью

1.111.ПИСАТЕЛ- 2.111.ПОДАРИ- 1.320.УЧЕНИК- 3.413.НОВ- 1.413.КНИГ- С 3.513.ДАРСТВЕНН-1.513.НАДПИС-

1.111.ПИСАТЕЛ- 2.111.ПОДАРИ,
3.413.НОВ- 1.413.КНИГ,
3.513.ДАРСТВЕНН-1.513.НАДПИС-;

ПОДАРИ- 3.УЧЕНИК- 4.КНИГ- С5.НАДПИС-

ПОДАРИ- 3.УЧЕНИК-,
ПОДАРИ- 4.КНИГ-,
ПОДАРИ- С5.НАДПИС-,
УЧЕНИК- 4.КНИГ-,
КНИГ- С5.НАДПИС-

ПОДАРИ- 4.КНИГ-
ПОДАРИ- 3.УЧЕНИК-
КНИГ- С5.НАДПИС-

⇒

ПОДАРИ- 3 / 4
КНИГ- С5,

ПОДАРИ- 4.КНИГ-,
КНИГ- С5.НАДПИС-

*Академия управления
при Президенте Республики
Беларусь*

Шибут М.С., Яковичин В.С.

РАСПОЗНАВАНИЕ
ГРАММАТИЧЕСКОЙ ИНФОРМАЦИИ
В ПРОЦЕССЕ ФОРМИРОВАНИЯ
ЛИНГВИСТИЧЕСКИХ ЗНАНИЙ

Спасибо за внимание

(+375 17) 220-48-46

m_shi@tut.by

7. Типы управления прилагательного

Тип управления	Объекты		Примеры
	o_1	o_2	
1	2	3	4
$q1$	1.G□		<i>Достойный уважения</i>
$q2$	1.D□	в 1.P□	<i>Покорный судьбе</i>
$q3$	1.I□		<i>Сильный духом</i>
$q4$		на 1.A□	<i>Лицом похож на мать</i>
$q5$	в 1.P□		<i>Сильный в арифметике</i>
$q6$	для 1.G□		<i>Интересный для всех</i>
$q7$	до 1.G□		<i>Жадный до денег</i>
$q8$	к 1.D□		<i>Склонный к размышлениям</i>
$q9$	перед 1.G□		<i>Беззащитный перед судьбой</i>
$q10$	с 1.I□		<i>Почтительный со всеми</i>

7. Типы управления

ОУЩОСТВИТЕЛЬНОГО

Тип управления	Объекты		Примеры
	O_1	O_2	
s_1	в $1.A \square$		<i>Превращение в сахар</i>
s_2	$1.I \square$		<i>Заражение энтузиазмом</i>
s_3	в $1.P \square$		<i>Превосходство в силе</i>
s_4	за $1.A \square$		<i>Оплата за труд</i>
s_5	к $1.D \square$		<i>Готовность к подвигу</i>
s_6	на $1.A \square$		<i>Реакция на свет</i>
s_7	перед $1.I \square$		<i>Восхищение перед картиной</i>
s_8	против $1.G \square$		<i>Борьба против войны</i>
s_9	с $1.I \square$		<i>Интервью с писателем</i>
s_{10}	$1.D \square$		<i>Гимн труду; памятник герою</i>
s_{11}		в $1.P \square$	<i>Помощь отцу в работе</i>
s_{12}	между $1.I \square$		<i>Дружба между народами</i>
s_{13}	на $1.A \square$		<i>Материал на костюм</i>
s_{14}	над $1.I \square$		<i>Победа над противником</i>
s_{15}	о $1.P \square$		<i>Отзыв о работе</i>

7. Типы управления глагола

Тип управления	Объекты		Примеры
	O ₁	O ₂	
v1	1.G□		<i>Добиваться успеха</i>
v2			<i>Радоваться удаче</i>
v3	1.D□	1.I□	<i>Грозить врагу войной</i>
v4		в 1.P□	<i>Подражать мастеру в работе</i>
v5		за 1.A□	<i>Заплатить автору за статью</i>
v6			<i>Видеть дом</i>
v7		1.G□	<i>Удостоить работу внимания</i>
v8		1.D□	<i>Вручить награды героям</i>
v9		1.I□	<i>Наполнить стакан водой</i>
v10	1.A□	в 1.A□	<i>Ввернуть лампочку в патрон</i>
v11		в 1.P□	<i>Изобличить свидетеля во лжи</i>
v12		за 1.A□	<i>Благодарить коллег за помощь</i>
v13		из 1.G□	<i>Извлечь корень из числа</i>
v14		к 1.G□	<i>Прикрепить ручку к двери</i>

7. Типы управления глагола-2

Тип управления	Объекты		Примеры
	o ₁	o ₂	
v15	1.A□	на 1.A□	<i>Вдохновить народ на подвиг</i>
v16		о 1.P□	<i>Знать правду о событиях</i>
v17		от 1.G□	<i>Беречь имущество от пожара</i>
v18		с 1.G□	<i>Взыскать налоги с населения</i>
v19		с 1.I□	<i>Поздравлять воинов с наградой</i>
v20		у 1.G□	<i>Просить внимания у слушателя</i>
v21		1.I□	
v22	без 1.G□		<i>Обходиться без помощи</i>
v23	в 1.A□		<i>Верить в удачу</i>
v24	за 1.A□		<i>Радоваться за товарища</i>
v25	за 1.I□		<i>Наблюдать за событиями</i>
v26	к 1.D□		<i>Готовиться к подвигу</i>
v27	на 1.A□		<i>Сердиться на соседей</i>
v28	над 1.I□		<i>Властвовать над миром</i>
v29	о 1.A□		<i>Споткнуться о порог</i>

7. Типы управления глагола-3

Тип управления	Объекты		Примеры
	O ₁	O ₂	
v30	о 1.P□		<i>Заботиться о детях</i>
v31	от 1.G□		<i>Воздержаться от суждений</i>
v32	перед 1.I□		<i>Благоговеть перед красотой</i>
v33	по 1.D□		<i>Тосковать по родине</i>
v34	против 1.G□		<i>Восстать против зла</i>
v35	с 1.P□		<i>Воевать с пережитками</i>