

Буторина Елена Петровна

СИСТЕМАТИЗАЦИЯ ИНСТИТУЦИОНИЗИРОВАННЫХ ЕДИНИЦ В СПРАВОЧНОЙ БАЗЕ ДАННЫХ

В статье обсуждаются принципы построения справочной базы данных институционализированных (статистически идиоматичных) неоднословных единиц, используемых в официальной сфере коммуникации. Рассматриваются единые формы представления разных видов институционализированных единиц и способы записи их дискурсивных и грамматических характеристик. Предлагаются способы выявления и перевода институционализированных единиц в текстах документов на разных языках.

Адрес статьи: www.gramota.net/materials/2/2016/3-1/23.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2016. № 3(57): в 2-х ч. Ч. 1. С. 85-87. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2016/3-1/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: phil@gramota.net

PHONETIC PECULIARITIES OF THE AMERICAN, CANADIAN, AUSTRALIAN, NEW ZEALAND VARIANTS OF THE ENGLISH LANGUAGE

Bodnar Svetlana Sergeevna, Ph. D. in Pedagogy
Kazan (Volga region) Federal University
ssr80@yandex.ru

The article examines the basic phonetic peculiarities of the American, Canadian, Australian, New Zealand variants of the English language. The interest for the problems of variability of the English language is motivated by the heterogeneity and complexity of its structure, specifics of its functioning in the different linguistic situations and territorial areas. The findings allow concluding on the level of influence of the variants under geographic closeness which is especially relevant nowadays when the English language becomes a part of language policy virtually in all the countries.

Key words and phrases: phonetic peculiarities; American variant of the English language; Canadian variant of the English language; Australian variant of the English language; New Zealand variant of the English language.

УДК 80:821.161.1

В статье обсуждаются принципы построения справочной базы данных институционализированных (статистически идиоматичных) неоднословных единиц, используемых в официальной сфере коммуникации. Рассматриваются единые формы представления разных видов институционализированных единиц и способы записи их дискурсивных и грамматических характеристик. Предлагаются способы выявления и перевода институционализированных единиц в текстах документов на разных языках.

Ключевые слова и фразы: категория официальности; база данных; институционализированные единицы; русский язык; перевод документов.

Буторина Елена Петровна, к. филол. н., доцент
Российский государственный гуманитарный университет
ep_butorina@il-rggu.ru

СИСТЕМАТИЗАЦИЯ ИНСТИТУЦИОНАЛИЗИРОВАННЫХ ЕДИНИЦ В СПРАВОЧНОЙ БАЗЕ ДАННЫХ

Активно используемые в текстах документов на разных языках институционализированные (статистически идиоматичные) единицы [1], образуют определённую систему. Такая система включает единицы разных типов: например, сочетание *В связи со сложной экономической ситуацией...* представляет собой клише, *Гарантируем возврат кредита в сумме...* – коллигацию, а *банковские реквизиты* – статистическую коллокацию.

Системное представление подобных единиц необходимо для решения целого ряда задач: это, например, составление документов, их перевод и лингводидактическое описание. Разноуровневые единицы различной степени сложности следует включать в общий справочник, материалы которого будут упорядочены с учётом интересов разных групп пользователей и должны содержать информацию не только об общезыковых свойствах таких единиц, но и о тех особенностях, которые подобные единицы демонстрируют в составе официальных документов.

Использование институционализированных единиц позволяет тратить минимальное количество времени на поиск оптимальных средств описания типизированных ситуаций в официальной сфере коммуникации. Так, клише зачастую используются для описания мотивации того или иного действия, например: *В порядке проведения совместной работы...; В подтверждение нашей договоренности...* и т.д. Коллигации нередко выступают как значимые компоненты текста документа, без которых он не обладает достаточной юридической силой. Структура коллигации в этом случае, как правило, включает явные или неявные перформативы, употребляемые во множественном числе (*Предлагаем...*). Возможны и сочетания разных видов институционализированных единиц: *Оплату гарантируем. Наши банковские реквизиты...* и т.д.

Для облегчения и упорядочения процессов составления документов на русском языке и их перевода на другие языки нами разработан формат справочника, реализованного как электронная база данных. В предлагаемом справочнике для институционализированных единиц, указываются жанры (виды) документов и соответствующие их разделы, в которых эти единицы могут употребляться с наибольшей вероятностью. Включение информации о разделе документа в словарную статью базы данных может облегчить задачу поиска нужной институционализированной единицы для анализа или синтеза текста определённого вида. Например, неоднословная единица *возложены (следующие) функции* с большой вероятностью может быть употреблена в таком разделе должностной инструкции, как «Функции». Поэтому для этой единицы в поле «Жанр/вид документа» указывается *должностная инструкция*, а в поле «Раздел документа» – *функции*.

Предполагается также указание грамматических особенностей, свойственных тем или иным единицам (коллокациям, клише и др.) в составе конкретного вида документа (например, должностной инструкции). Для некоторых слов, например, могут указываться только те формы, которые функционируют в этих видах

документов, а не в языке вообще. Так, например, глагол *возложить* (*возложить обязанности*) имеет в языке помимо инфинитива спрягаемые формы, а также формы причастий и деепричастия, однако если в ходе исследования текстов корпуса соответствующих документов он встретился только в формах инфинитива и страдательного причастия прошедшего времени (например, в текстах должностных инструкций), то в разделе базы данных для такого вида документов, как должностные инструкции, он будет приведен именно в этих формах.

Некоторые из компонентов одного несвободного сочетания, представляющего собой институционализированную единицу, могут входить в состав другого, как например, *опыт работы* в составе сочетания *опыт [стаж] (Adj₂) работы (Prep (Adj₆) N_{3/6}) от [не менее] + Num_{card} N₂)* (см. об этом подробнее в [2], там же приводится список работ, в которых обсуждаются подобные формы записи). Приведенная форма записи предложена И. С. Мироновой [Там же] и может быть реализована в электронном справочнике в виде гипертекста с последующими элементами автоматизации выбора институционализированных единиц из соответствующего перечня базы данных для заполнения переменных позиций в конкретном документе. Для такой работы представляется наиболее удобным формат электронной базы данных, а не обычного словаря, так как справочник такого типа должен включать помимо собственно языковых и другие коммуникативные единицы, используемые в официальной сфере коммуникации. Рассмотрим форму предлагаемого представления институционализированных единиц более подробно.

Неизменные (опорные) компоненты институционализированных единиц выражаются во всех документах одним и тем же словом (словосочетанием), в таком же виде они будут приводиться и в справочнике (например, *опыт... работы* в приведённом ранее примере). Переменные компоненты обозначаются не конкретными словами, а наименованием той части речи, к которой относится переменная. Принадлежность к той или иной части речи маркируется при помощи традиционных условных наименований: *N*, *Adj*, *V*, *Adv*, *Num*, *Pron* и т.п. Используются обозначения и для отдельных подклассов тех или иных частей речи: например, *Num card* – количественное числительное. Это касается и форм – так, падежные формы обозначаются при помощи числительных: 1 – *Им.п.*; 2 – *Род.п.*; 3 – *Дат.п.* и т.п. Например, обозначение *Adj₂* используется для прилагательного в родительном падеже. В квадратных скобках может указываться менее вероятный вариант или синоним опорного слова: например, *опыт [стаж]*. В круглые скобки заключены факультативные для заполнения позиции, например: *(Adj₂) работы*. В некоторых случаях факультативные компоненты следуют друг за другом, при этом каждый из них записывается в том порядке, в каком они встречаются в тексте, и заключается в отдельные скобки. Знаком «+» обозначены границы опорных компонентов. Если существует возможность прономинализации, то заменяемый и заменяющий компоненты обозначаются следующим образом: <компонент> (на время <Pron₂> отсутствия + <N₂>).

Существуют институционализированные единицы, компоненты которых могут быть разделены другими словами (например, *опыт... работы*). В словарных статьях разделяющие слова будут указаны в круглых скобках; при этом если разбивать институционализированную единицу в одной синтаксической позиции потенциально могут слова, относящиеся к разным частям речи, то они будут разделены знаком «/» и вместе заключены в одни скобки, а первым будет стоять то из слов, которое чаще остальных встречается в этой позиции. В тех случаях, когда слова разных частей речи могут одновременно присутствовать между элементами институционализированной единицы, каждое из них будет указано отдельно в скобках в том порядке, в котором они встречаются в тексте: например, *исполнение (своих) (должностных) обязанностей*. При каждой институционализированной единице в скобках будет указано, является ли она «незаменимой» (не допускающей замены компонентов другими единицами – помета *нз*) или «неразрывной» (не допускающей вставки дополнительных слов между компонентами – помета *нр*). Следует уточнить, что анализ этих свойств рассматриваемых единиц производится с точки зрения конкретных видов документов. Поэтому если один из компонентов может быть заменен только словом, не относящимся к институционализированным единицам официальной сферы коммуникации, такие сочетания будут помечены как *нз*. Например, существительное *закрепление* в единице *закрепление кадров* теоретически могло бы быть заменено на **удержание* или **фиксация*, а прилагательное *настоящая* в клише *настоящая инструкция* – на местоимение **эта*, однако такие сочетания не используются в официальной сфере коммуникации.

В предлагаемый справочник предполагается включить информацию о наиболее вероятной связи институционализированных единиц с другими элементами предложения и текста. Подобная информация оформляется в виде падежных вопросов от институционализированной единицы к связанным с ней словам за её пределами. Подобные вопросы представляют собой обязательные валентности рассматриваемых единиц. Например, словосочетание *построить взаимоотношения* будет сопровождаться вопросами: *в/на чем?*; *между кем и кем?* и образцами из документов, содержащих примеры практической реализации приведенных в словарной статье (разделе базы данных) зависимостей: *построить взаимоотношения в коллективе / на предприятии / между работником и работодателем*.

В соответствующих разделах базы данных для институционализированных единиц фиксируются только те их свойства, которые встретились в корпусе определённого вида текстов. Так, в единице *с (Adj₃) N₅ означен(а)* для текстов должностных инструкций переменный компонент (*N₅*) выражается только существительным *инструкция* в творительном падеже. Теоретически он может выражаться в русском языке любым существительным, обозначающим текст, однако в корпусе текстов, например, должностных инструкций не было выявлено ни одного случая употребления других существительных в соответствующей позиции рассматриваемой институционализированной единицы. Поэтому в качестве конкретного примера формы

выражения переменного компонента этой единицы, в разделе для должностных инструкций будет указано только одно существительное – *инструкция*.

Предложенная форма представления институционализированных единиц [Там же] апробирована на материале русского и итальянского языков. Использование единой формы записи позволяет выявить специфические компоненты в документах на разных языках, обусловленные традициями соответствующей деловой культуры. Подобные справочные материалы позволили бы существенно упростить и унифицировать перевод разных видов документов.

Помимо единой формы записи институционализированных единиц могут быть использованы единые принципы их статистического выявления в текстах на разных языках (см. например, [3; 4]). К. В. Соловьёва предложила поэтапную процедуру перевода институционализированной единицы на другие языки. Словарная статья для этого делится на три зоны:

«1) Зона 1: одна (свободная) лексема (сюда помещаются однословные эквиваленты перевода);

2) Зона 2: неоднословная номинация (в данную зону попадают сочетания, ядром которых является данная лексема);

3) Зона 3: идиома (здесь будут находиться идиоматические выражения, в которые входит лексема).

При этом каждая зона получает определенный приоритет (weight). Например, Зона 1 (weight: 10), Зона 2 (weight: 100), Зона 3 (weight: 1000). Таким образом, программа находит некоторую лексему, “захватывая” левый и правый контекст (–5 лексема +5) и направляется в словарь. Сначала проверяется зона с наиболее высоким приоритетом (Зона 3) – если в ней существует сочетание, равное данной лексеме и контексту / части контекста, выбирается перевод данного сочетания и отправляется на “выход” (при этом, контекст, найденный в словаре, маркируется как переведенный и вошедшие в него лексемы далее отдельно не проверяются по словарным статьям). Если такого сочетания нет, программа переходит в зону с меньшим приоритетом (Зона 2) и продельвает аналогичную операцию. Если на этом уровне также не находится эквивалента, для перевода выбирается лексема из Зоны 1 (рандомно)» [4, с. 265].

Некоторые документы можно не переводить, а сразу составлять на нескольких языках, если существует единообразное описание коммуникативной категории официальности и её единиц для разных языков и деловых культур.

Таким образом, представляются правомерными следующие выводы.

1. В качестве справочника, включающего используемые в официальной сфере взаимодействия стандартные коммуникативные единицы разной природы, может быть предложена единая база данных.

2. Формирование такой базы данных предполагает включение в неё институционализированных (статистически идиоматичных) неоднословных единиц разной степени сложности (коллокации, клише и др.).

3. Для единообразного представления институционализированных единиц может быть предложена унифицированная форма их записи.

4. Принципы выявления таких единиц в документах на разных языках тоже могут быть едиными и представлять собой статистические процедуры.

5. Справочная база данных может включать помимо институционализированных единиц, используемых в официальной коммуникации на русском языке, эквивалентные им единицы на других языках. Выявление подобных соответствий и их единообразное представление может значительно облегчить решение таких задач, как составление документов, их перевод и лингводидактическое описание.

Список литературы

1. **Буторина Е. П.** Институционализированные единицы в официальном дискурсе и медиа // Филологические науки. Вопросы теории и практики. Тамбов: Грамота, 2015. № 12 (54). Ч. 1. С. 54-56.
2. **Буторина Е. П., Миронова И. С.** Культурно обусловленное знание в инструктивных текстах [Электронный ресурс] // Современные исследования социальных проблем (электронный научный журнал). № 9 (29). URL: <http://journals.org/index.php/sisp/article/view/9201369> (дата обращения: 11.08.2015).
3. **Буторина Е. П., Соловьёва К. В.** Выявление несвободных сочетаний слов как основы институционального дискурса // Международный аспирантский вестник. Русский язык за рубежом. М., 2012. № 2. С. 11-14.
4. **Буторина Е. П., Соловьёва К. В.** О выделении неоднословных номинаций в юридических текстах // European Social Science Journal (Европейский журнал социальных наук). М., 2012. № 9 (2). С. 252-267.

SYSTEMATIZATION OF INSTITUTIONALIZED UNITS IN THE REFERENTIAL DATABASE

Butorina Elena Petrovna, Ph. D. in Philology, Associate Professor
Russian State University for the Humanities
ep_butorina@il-rsggu.ru

The article discusses the principles to develop a referential database of the institutionalized (statistically idiomatic) multi-word units used in the official communication. The author examines the unified forms to represent the different types of institutionalized units and the methods to record their discursive and grammatical characteristics. The paper introduces the means to identify and translate the institutionalized units in the documents in different languages.

Key words and phrases: category of officialism; database; institutionalized units; Russian language; translation of documents.