

*А.А. Виландеберг*

**КОРПУС ПАРАЛЛЕЛЬНЫХ ПРАВОВЫХ  
ДОКУМЕНТОВ КАК СОСТАВНАЯ ЧАСТЬ  
АРМ ЮРИСТА-ПЕРЕВОДЧИКА**

Процесс становления правового государства в России, построение демократического общества, расширение международного сотрудничества привели к увеличению количества источников нормативно-правовой информации, пользование которой является процессом, требующим больших временных и трудовых затрат. Возникла реальная ситуация активного обмена информацией и необходимость анализа и обработки огромного потока этой информации на разных языках<sup>1</sup>. Сегодня привлечение информации на разных языках в конкретной предметной области права становится практически необходимым для любого юриста, особенно для специалиста, занимающегося вопросами международного права, а также вопросами защиты прав человека.

Анализ различных существующих АРМ, автоматизированных информационно-поисковых правовых систем и отдельных систем обработки текста, разнообразие словарей и тезаурусов позволяет сделать вывод о необходимости создания системы, ориентированной на потребности специалистов различных предметных областей знаний, а в данном конкретном случае на потребности специалистов, занимающихся вопросами соблюдения и защиты прав ребенка на внутреннем и международном уровнях. При этом следует учитывать, что вопросами защиты прав ребенка могут заниматься не только специалисты-юристы, имеющие соответствующее образование, но и учителя, врачи, родители, а также люди, занимающиеся этими вопросами на добровольной основе. Именно поэтому разнообразие терминологических сис-

---

<sup>1</sup> *Беляева Л.Н.* Лингвистические автоматы в современных информационных технологиях. СПб., 2001. С. 121.

тем и несоответствие юридических концепций в различных странах порой приводит к непониманию конкретных правовых ситуаций и возможным нарушениям прав ребенка. Следовательно, возникает вопрос гармонизации правовой терминологии.

АРМ юриста и юриста-переводчика должно обеспечить специалиста возможностью собирать, хранить, искать и вкладывать информацию на естественном языке по той или иной правовой тематике, соблюдать формальные характеристики правовых документов, а также выполнять информационные запросы для поиска требуемых юридических документов. АРМ юриста представляет собой набор средств правовой коммуникации, т.е. набор представлений, приемов передачи и получения правовой информации.

Автоматизированное рабочее место специалиста-юриста, переводчика в узко-предметной области знаний несомненно должно включать различные информационно-поисковые системы, а также двуязычные (многоязычные) автоматические словари, тезаурусы и глоссарии. Создание АРМ юриста позволит добиться унификации терминологии конкретных областей права, что особенно важно при переводе государственных документов, законодательных актов, правительственных постановлений и директив, а также иных юридических документов по защите прав человека. При этом следует отметить, что в рамках создания АРМ юриста нас будет интересовать терминология права, которая, в отличие от терминологии юриспруденции, является терминологией правоприменительной практики.

База данных по правовой области знаний, входящая в АРМ юриста-переводчика, должна включать:

- полные тексты необходимых правовых документов (корпус параллельных текстов ООН по правам ребенка). Включение в состав АРМ специалиста в области защиты прав человека корпуса параллельных текстов основополагающих документов ООН дает возможность ис-

пользовать указанный корпус не только для перевода вновь создаваемых документов, но и как своеобразную справочную базу данных. В данный корпус представляется возможным также включить международные договоры, обеспечивающие общую защиту прав человека.

- законодательные документы, касающиеся соблюдения прав ребенка. В РФ это могут быть – Конституция РФ, Кодекс о браке и семье РФ, Гражданский кодекс РФ, Уголовный кодекс РФ, Закон об образовании и др.;
- требования к оформлению различных документов, представленные в виде шаблонов.

При создании АРМ юриста-переводчика и определении задач его функционирования, следует в первую очередь учитывать такие специфические черты языка права, как точность, однозначность, интегрированность, унифицированность, обобщенность, формализованность правовой терминологии.

Следовательно, центральным вопросом создания информационно-поискового языка законодательства является терминология.

Проблема гармонизации терминов представляется особенно важной в области защиты и соблюдения прав ребенка. Вопросы защиты прав ребенка решаются, согласно государственному законодательству, по-разному в каждой отдельно взятой стране, при основополагающей роли документов ООН<sup>1</sup> по защите прав ре-

---

<sup>1</sup> *Convention on the Rights of the Child – Конвенция о правах ребенка (1990), Declaration on Social and Legal Principles relating to the Protection and Welfare of Children, with special reference to Foster Placement and Adoption Nationally and Internationally – Декларация о социальных и правовых принципах, касающихся защиты и благополучия детей, особенно при передаче на воспитание и усыновление на национальном и международном уровнях (1986), United Nations Rules for the Protection of Juveniles Deprived of their Liberty – Руководящие принципы ООН для предупреждения преступности среди малолетних (1990), United Nations Rules for the Protection of Juveniles Deprived of their Liberty – Правила*

банка. Учитывая все более возрастающее число контактов между государствами, различными организациями и отдельными гражданами на международном уровне, иногда приводящих к довольно конфликтным ситуациям и заканчивающихся решениями суда, необходимо четко гармонизировать термины данной предметной области знаний и сделать полученные результаты доступными и, что не менее важно, понятными для населения в целом (например, *усыновление/передача на воспитание, обязанности родителей/обязанности опекунов/попечителей*).

Терминологические двуязычные словари предметной области соблюдения и защиты прав ребенка являются подструктурой терминологического банка данных и могут являться частью АРМ юриста-переводчика. В связи с тем, что термины являются идеальным средством структурирования научного знания, терминологический банк данных может стать наилучшим способом организованного хранения представленной информации.

Нам представляется совершенно очевидным, что основным материалом, который может быть использован при создании АРМ юриста, занимающегося вопросами соблюдения и защиты прав ребенка, являются разнообразные документы по данной правовой тематике международного и региональных уровней. Среди таких документов в первую очередь необходимо выделить документы ООН, как признанного мирового лидера в вопросах защиты прав человека.

Все правовые документы ООН составляются сразу на шести официальных языках, при этом документы имеют одинаковую юридическую силу, т.е. считаются одинаково подлинными и равнозначными, или аутентичными. Эти же документы можно определить и как корпус абсолютно параллельных текстов, т.е. текстов, не различающих текст оригинала и текст перевода.

---

*ООН, касающиеся защиты несовершеннолетних, лишенных свободы (1990).*

Так в настоящем исследовании в корпус текстов были включены документы ООН, касающиеся соблюдения и защиты прав ребенка. В перспективе в корпус будут добавлены документы ООН, касающиеся защиты женщин и детей в период вооруженных конфликтов, что представляется достаточно актуальным. Соблюдая требования, предъявляемые к корпусу параллельных текстов, мы получили стандартный корпус многократного использования<sup>1</sup>. Представленный в исследовании корпус текстов

---

<sup>1</sup> Согласно спецификации проекта EAGLES  
(<http://www.ilc.pi.cnr.it/EAGLES96/corpus/corpus.html>)  
этот корпус имеет следующие характеристики:

- **по модусу, форме существования** – письменные аутентичные тексты;
- **по типу текстов** – исследуемые тексты являются официальными документами ООН, имеющими одинаковую юридическую силу;
- **по объему** – небольшой подкорпус;
- **по языковому покрытию** – корпус подъязыка права; достаточно представительный по количеству словоупотреблений (52766);
- **по жанру (регистру)** – юридические тексты;
- **по варьированию языков** – параллельный переводной двуязычный;
- **по общности, которая продуцирует корпус** – предназначен для носителей русского и английского языков, занимающихся вопросами соблюдения и защиты прав ребенка;
- **по маркированности** – представленные тексты являются одинаково подлинными и равнозначными (аутентичными), т.е. представляют собой корпус абсолютно параллельных текстов, в котором не различаются текст оригинала и текст перевода;
- **по степени открытости корпуса** – существует возможность постоянного пополнения и накопления данного подкорпуса;
- **по национальному варьированию** – это часть корпусов русского и английского языков, с возможной перспективой дополнения корпуса соответствующими параллельными документами ООН на остальных официальных языках ООН.

относится к понятийной области юриспруденции, а именно защиты прав ребенка, и отражает правовую терминологию.

Исходя из того, что каждый текст в корпусе является исходным, можно говорить о том, что в исследуемом корпусе употребляются термины и устойчивые терминологические сочетания, зафиксированные в данных языках и понятные рядовому носителю языка. Следовательно, эти термины являются стандартизированными, рекомендуемыми и устоявшимися<sup>1</sup>.

В связи с этим корпус параллельных текстов представляется наиболее удобным для лингвистического анализа, в том числе для составления терминологических баз данных, частотных словарей, тезаурусов. Ведь при моделировании терминологии следует исходить из того, что юридическая терминология является основным, наиболее информативным пластом лексики языка законодательства, способствующим точному и ясному формулированию правовых предписаний.

Процедура анализа корпуса параллельных текстов состоит из нескольких этапов.

1. Фрагментация исходных текстов.
2. Выравнивание текстов на уровне абзацев и предложений (text alignment) или параллельная сегментация текстов<sup>2</sup>.
3. Выравнивание фрагментов на уровне слов и словосочетаний (word alignment).

- 
- **по историческому варьированию** – синхронный подкорпус;
  - **по возрасту пользователя** – фиксирует официальные языки, предназначенные, в основном, для понимания взрослыми;
  - **по доступности подкорпуса** – в перспективе этот корпус будет свободно доступен в Интернете для изучения и использования в переводческих целях.

<sup>1</sup> *Беляева Л.Н., Герд А.С., Убин И.И.* Автоматизация и лексикография // Прикладное языкознание / Отв. ред. А.С. Герд. СПб., 1996. С. 319.

<sup>2</sup> *Беляева Л.Н.* Лингвистические автоматы в современных информационных технологиях. СПб., 2001. С. 63.

4. Создание частотного и алфавитно-частотного словарей каждого отдельно взятого корпуса.

5. Создание конкорданса каждого отдельно взятого корпуса.

6. Рабочий машинный перевод (англо-русский) полученных словарей и текстов.

7. Сопоставительный лексико-семантический и морфологический анализ полученных результатов.

Рассмотрим указанные этапы.

1. На начальном этапе происходит процесс фрагментации исходных текстов, а именно пофрагментное сопоставление текстов на английском (массив 1) и русском (массив 2) языках, производящееся на уровне предложений.

*Фрагментация исходных текстов*

<i>Part I Article 1 For the purposes of the present Convention, a child means every human being below the age of eighteen years unless, under the law applicable to the child, majority is attained earlier.</i>	<i>Часть I Статья 1 Для целей настоящей Конвенции ребенком является каждое человеческое существо до достижения 18-летнего возраста, если по закону, применимому к данному ребенку, он не достигает совершеннолетия ранее.</i>
--	---

2. На втором этапе производится процедура двуязычного выравнивания, которая использует два параллельных корпуса текстов и устанавливает, какое предложение из массива 1 соответствует предложению из массива 2. Предварительно выровненный корпус используется далее как справочная база данных<sup>1</sup>.

---

<sup>1</sup> *Беляева Л.Н.* Проблемы анализа корпусов параллельных текстов с различной графикой // Доклады научной конференции «Корпусная лингвистика и лингвистические базы данных». СПб., 2002. С. CCC.

3. Одновременно с выравниванием параллельных текстов на уровне абзацев и предложений (text aligning) проводится сопоставление на уровне слов и словосочетаний (word aligning). Выравнивание корпусов на уровне слов дает возможность получения списков самих переводных эквивалентов (ПЭ). Задача поиска ПЭ предполагает получение в конечном итоге соответствия для каждого термина или терминосочетания в тексте.

*Выравнивание текстов на уровне сопоставимых фрагментов*

<i>For the purposes of the present Convention, a child means every human being below the age of eighteen years unless, under the law applicable to the child, majority is attained earlier.</i>	<i>Для целей настоящей Конвенции ребенком является каждое человеческое существо до достижения 18-летнего возраста, если по закону, применимому к данному ребенку, он не достигает совершеннолетия ранее.</i>
---	--

С точки зрения терминологического анализа, сегментация предложений на уровне слов и словосочетаний позволяет выделить в корпусе параллельных текстов термины и устойчивые терминологические словосочетания (правовые клише) и их переводческие эквиваленты.

*Терминологические словосочетания и их переводческие эквиваленты*

<i>administration of juvenile justice</i>	<i>отправление правосудия в отношении несовершеннолетних</i>
<i>adoption nationally and internationally</i>	<i>усыновление на национальном и международном уровнях</i>
<i>consistent with human rights and fundamental freedoms</i>	<i>в соответствии с правами человека и основными свободами</i>
<i>if committed by a young person</i>	<i>в случае совершения молодыми лицами</i>

<i>illicit transfer and non-return of children abroad</i>	<i>незаконное перемещение и невозвращение детей из-за границы</i>
<i>national child welfare services</i>	<i>национальные службы, занимающиеся заботой о детях</i>

4. Создание частотных словарей позволяет выявить определенные закономерности словоупотреблений в каждом отдельном массиве текстов, сравнить их между собой и выявить группу слов и словосочетаний, наиболее часто употребляющихся в обоих массивах.

По информации о частотности словоупотреблений, взаимной частотности в пределах предложения, текста и корпуса в целом можно сделать вывод о том, насколько часто данные комбинации слов встречаются в подязыке данной ПО и предложить наиболее соответствующие переводческие эквиваленты.

5. Создание конкорданса данного корпуса и включение его в состав АРМ юриста-переводчика, позволит переводчику выбирать примеры на нужное слово из данного корпуса текстов в автоматическом режиме, сокращая время, затрачиваемое на перевод правовых документов.

<b><i>adoption</i></b>	<b><i>усыновление</i></b>
<i>adopt</i>	<i>усыновить</i>
<i>adopted</i>	<i>усыновленный</i>
<i>adoption is permissible</i>	<i>усыновление допустимо</i>
<i>adoption of a child</i>	<i>усыновление ребенка</i>
<i>adoption placements</i>	<i>условия усыновления</i>
<i>child to be adopted</i>	<i>усыновляемый ребенок</i>
<i>foster or adoptive family</i>	<i>приемная семья или семья, берущая ребенка на воспитание</i>
<i>foster or an adopted child</i>	<i>усыновленный ребенок или ребенок, взятый на воспитание</i>
<i>in respect of national adoption</i>	<i>в соответствии с законами усыновления внутри страны</i>
<i>intercountry adoption</i>	<i>усыновление за границей</i>

<i>legal validity of the adoption</i>	<i>законность усыновления</i>
<i>member of the adoptive family</i>	<i>члены приемной семьи</i>
<i>national adoption</i>	<i>усыновление внутри страны</i>
<i>primary aim of adoption</i>	<i>основная цель усыновления</i>
<i>prior to the adoption</i>	<i>перед усыновлением</i>
<i>prospective adoptive parents</i>	<i>будущие приемные родители</i>
<i>system of adoption</i>	<i>системы усыновления</i>

6. Экспериментальный машинный перевод полученных словарей позволяет выделить следующие группы терминов в исследуемом корпусе текстов:

- слова и специфические словосочетания, не требующие дополнительной обработки (полные соответствия за счет информации из АС и синтаксического анализа):

<i>массив 1</i>	<i>машинный перевод</i>	<i>массив 2</i>
<i>disciplinary requirements and procedures</i>	<i>дисциплинарные требования и процедуры</i>	<i>дисциплинарные требования и процедуры</i>
<i>responsibility for the upbringing and development of the child</i>	<i>ответственность за воспитание и развитие ребенка</i>	<i>ответственность за воспитание и развитие ребенка</i>

- словосочетания, не опознанные системой перевода и требующие ввода дополнительных параметров в систему:

<i>массив 1</i>	<i>машинный перевод</i>	<i>массив 2</i>
<i>prevention of juvenile delinquency</i>	<i>предотвращение юношеского невыполнения обязанностей</i>	<i>предупреждение преступности среди несовершеннолетних</i>
<i>staff trained in preventive health care and the handling of medical emergencies</i>	<i>штат служащих обучился профилактическому здравоохранению и обработке медицинского критического положения</i>	<i>персонал, обученный оказывать профилактический уход и неотложную медицинскую помощь</i>

или корректировку уже существующих параметров:

<i>массив 1</i>	<i>машинный перевод</i>	<i>массив 2</i>
<i>every human being below the age of eighteen years</i>	<i>каждый человек ниже возраста восемнадцати лет</i>	<i>каждое человеческое существо до достижения 18-летнего возраста</i>
<i>nationality of the child differs from that of the prospective adoptive parents</i>	<i>национальность ребенка отличается от этого предполагаемых адоптивных родителей</i>	<i>гражданство ребенка иное, чем гражданство будущих приемных родителей</i>

Следовательно, одной из важнейших задач, возникающих при создании электронного словаря, является определение соответствующих параметров для ввода термина и его характеристик.

6. В аспекте гармонизации терминологии требуется выполнение процедуры лексико-семантического анализа полученных терминов и терминосочетаний.

Прежде всего, мы расклассифицировали термины по названию реалий данной проблемной области.

- Названия международных организаций и документов.

Следует учитывать, что употребление этих названий в текстах переводов, особенно правовых документов, требует принятия стандартного переводческого решения. В связи с этим, на основе анализа корпуса параллельных правовых текстов необходимо выделение названий международных организаций, их четкая унификация и внесение отдельным пунктом в состав базы данных АРМ юриста-переводчика.

- Юридические термины латинского происхождения.

Проведенный анализ корпуса параллельных текстов показал, что в данной терминосистеме употребление латинизмов не является распространенным. Это объясняется необходимостью доступности терминологии данной предметной области для боль-

шого числа населения. В исследуемом корпусе было выявлено всего лишь 4 термина, заимствованных из латинского языка, довольно прочно вошедшие в общеупотребительный язык.

<i>legislation</i>	<i>законодательство</i>
<i>sanctions</i>	<i>санкции</i>
<i>status</i>	<i>статус</i>
<i>juvenile justice system</i>	<i>система правосудия в отношении несовершеннолетних/ система ювенального права</i>

- Специальные термины общеюридического характера.

<i>as a measure of last resort</i>	<i>в качестве крайней меры</i>
<i>adequate counselling</i>	<i>надлежащие консультации</i>
<i>life imprisonment</i>	<i>пожизненное тюремное заключение</i>
<i>mechanisms for the appropriate coordination of prevention efforts</i>	<i>механизмы для надлежащей координации деятельности в области предупреждения преступности</i>

- Узкоспециальные термины ювенальной юстиции (детского права).

<i>abduction</i>	<i>похищение детей</i>
<i>concerned with the welfare of children</i>	<i>занимающийся вопросами благополучия детей</i>
<i>conditions under which a juvenile is detained</i>	<i>условия, в которых содержится под стражей несовершеннолетний</i>
<i>legal validity of adoption</i>	<i>юридическая сила усыновления.</i>

- Общеупотребительные слова в функции юридического термина.

<i>access</i>	<i>доступ</i>
<i>material assistance</i>	<i>материальная помощь</i>

Следует отметить, что использование слова общеупотребительной лексики (терминологизация) является отличительной чертой терминологии права. Так, связь терминов с лексикой общего употребления может быть различной:

- 1) термины распространяются так широко, что уже не воспринимаются как термины: *ребенок*;
- 2) термины носят специальный характер, но известны широкому кругу населения: *адвокат*;
- 3) термины известны лишь специалистам: *омбудсмен*.

Производимый компонентный анализ полученных терминосо сочетаний позволяет решить проблему многозначного юридического термина.

<i>administration of any drug</i>	<i>применение любого лекарственного средства</i>
<i>administration of juvenile justice</i>	<i>отправление правосудия в отношении несовершеннолетних</i>
<i>administration of the detention facilities</i>	<i>администрация исправительных учреждений</i>

Анализ параллельных текстов позволяет выявить разницу между словарным эквивалентом и переводным эквивалентом. Определение переводческих эквивалентов при анализе корпуса именно параллельных текстов способствует выявлению переводческих соответствий, зафиксированных в официальных правовых документах. Это дает нам возможность утверждать правомочность перевода как с языка А на язык Б, так и наоборот.

Последовательное выполнение всех вышеназванных операций позволяет проанализировать не только сами термины, терминологические словосочетания, но и устойчивые словосочетания с сохранением связей между словами, что особенно важно при создании электронного двуязычного словаря.

Исходя из вышесказанного, можно сделать вывод о том, что именно массивы абсолютно параллельных текстов могут быть использованы в качестве основного источника для создания дву-

язычных электронных словарей и баз данных в конкретных предметных областях знаний.

Помимо этого, представление корпуса параллельных текстов как составной части АРМ специалиста позволит значительно облегчить работу переводчика путем введения сегментированных фрагментов текстов в базу данных, что позволит переводчику в любой момент найти необходимую информацию не просто в виде отдельных лексических единиц, но и в виде словосочетаний, предложений, абзацев.