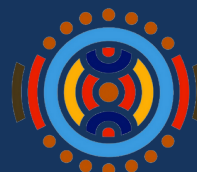


DE ZÉRO À NUMÉRISÉE SERIES:

# DIRECTIVES DE TERMINOLOGIE

РУКОВОДСТВО ПО ЦИФРОВИЗАЦИИ ЯЗЫКА

TRANSLATION  
COMMONS



2022-2032 | INTERNATIONAL DECADE OF

**Indigenous Languages**

Translation Commons © 2022

Настоящий материал распространяется на условиях международной лицензии Creative Commons «С указанием авторства» версии 4.0.

## Руководство по терминологии

Авторы: Сью Эллен Райт (Sue Ellen Wright),  
Акил Айер (Akil Iyer), Шуго Като (Shuto Kato)  
Рецензенты и соавторы: Крэйг Корнелиус  
(Craig Cornelius), Алайна Брандт (Alaina Brandt),  
Джули Андерсон (Julie Anderson),  
Текс Тексин (Tex Texin)  
Маркетинг: Леонидас Паппас (Leonidas Pappas)  
Координатор проекта: Паула Сирило (Paula Cirilo)

Мы будем рады предложениям по улучшению  
руководств.

Свяжитесь с нами: [krista@translationcommons.org](mailto:krista@translationcommons.org)  
Настоящий материал распространяется на условиях  
международной лицензии Creative Commons  
«С указанием авторства» версии 4.0.  
<https://creativecommons.org/licenses/by/4.0/>

1. 1. ПРЕДИСЛОВИЕ	4
2. 2. О ДАННОМ ДОКУМЕНТЕ	5
3. 3. ЧТО ТАКОЕ ТЕРМИНОЛОГИЯ И ТЕРМИНОЛОГИЧЕСКАЯ РАБОТА?	7
4. 4. ИДЕАЛЬНЫЙ РАБОЧИЙ ПРОЦЕСС	8
4.1 Вы готовы начать работу над терминологией?	9
4.2 Вы знаете, кто будет выполнять эту работу?	9
4.3 Создавайте тексты и собирайте существующие термины на вашем языке.	10
4.4. Придумывайте термины для описания новых идей и понятий.	11
4.5 Фиксируйте и актуализируйте терминологию в общих глоссариях.	11
5. ПРИМЕНЕНИЕ УКАЗАНИЙ РУКОВОДСТВА	12
5.1 Как пользоваться блок-схемой	12
5.2 Лексикография (составление словарей известных понятий) и терминология (терминография)	15
5.3 Как определяется терминология, и почему эта работа так важна?	16
6. ДОКУМЕНТИРОВАНИЕ ТЕРМИНОВ	20
7. ПРОЧИЕ РАБОТЫ	25
7.1 Создание корпуса текстов	25
7.2 Добавление переводов для создания корпуса параллельных текстов	26
7.3 Создание систем общих понятий	27
7.4 Системы понятий «часть-целое»	28
8. СОВМЕСТНОЕ ИСПОЛЬЗОВАНИЕ И ПУБЛИКАЦИЯ ТЕРМИНОЛОГИЧЕСКИХ РЕСУРСОВ	29
<b>ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ</b>	<b>31</b>

## 1. 1. ПРЕДИСЛОВИЕ

[Translation Commons](#) — некоммерческое сообщество волонтеров, которое занимается поддержкой преобразования языков в цифровой формат, курирует представителей языковых профессий, а также предоставляет образовательные и информационные ресурсы в сфере языковых услуг.

Одной из центральных программ Translation Commons является «Инициатива по цифровизации языков», которая дает заинтересованным языковым сообществам возможность получить доступ к цифровым технологиям. В мире существует около 6000 языков с незначительным присутствием в цифровом пространстве или вовсе в нем не представленных. «Инициатива по цифровизации языков» предлагает стратегический план, следуя которому любое сообщество сможет осуществить цифровизацию своего языка.

Для привлечения внимания к коренным народам и вопросам цифровизации их языков Translation Commons сотрудничает с ЮНЕСКО в рамках инициативы [«2019: Международный год языков коренных народов»](#). Часть миссии «Инициативы по цифровизации языков» — обеспечить доступность языков коренных сообществ и национальных меньшинств в цифровом формате, чтобы такие языковые сообщества могли участвовать во всемирной сетевой деятельности и пользоваться всеми преимуществами современных компьютерных приложений на своем родном языке. Такие руководства обеспечивают сообщества инструментами и знаниями для оцифровки письма и внедрению своих языков в Интернет, и соответственно, получить необходимую поддержку в этом процессе, не теряя своей независимости. Помимо создания руководств Translation Commons предоставляет обучающие материалы, проводит семинары и помогает языковым сообществам установить рабочие связи со специалистами отрасли, которые способны направить их в процессе стандартизации.

Данный документ входит в серию методических руководств под названием «На пути в цифровое пространство», в которой комплексно рассматривается деятельность по цифровизации языков. Авторами этих руководств выступили специалисты в сфере языковых технологий и лингвистики. Этот документ предназначен для любого языкового сообщества, которое заинтересовано в возможном применении родного языка в цифровых системах.

Присутствие языка в сети Интернет открывает новые каналы общения для его носителей. В разделе [«Преимущества цифровизации языков»](#) подробно рассматривается, каким образом от присутствия языка в цифровом пространстве выигрывает как коренной народ-носитель языка, так и мировое сообщество в целом.

Более подробная информация о процессе цифровизации языка приведена в разделе [«На пути в цифровое пространство. Как привести ваш язык в Интернет»](#). В разделе [Resources](#) (Ресурсы) вебсайта Translation Commons приведена дополнительная информация по теме Language Digitization (Цифровизация языков) и материалы: руководства, презентации, видео и прочие документы.

## 2. 2. О ДАННОМ ДОКУМЕНТЕ

В данном документе описан порядок фиксирования терминологии существующего, но недокументированного или недостаточно документированного языка. В нем описано, как записывать существующие слова и создавать новую терминологию для общения на специфические темы, для которых в языках коренных народов может не существовать специальных слов. Эта методология носит рекомендательный характер и представляет собой руководство по различным рабочим процессам по созданию или обновлению терминологии, включая планирование, сбор ресурсов, проведение терминологического анализа и документирование. Благодаря этому процессу коренные народы смогут вводить

термины по мере необходимости, например, в области технологий, медицины, политики, права или в других областях, которые могут быть за пределами исторической сферы их культуры или языка. Расширение и обновление такого лексического запаса поможет коренным народам получить доступ и использовать ресурсы, способствующие их включению в цифровую среду.

Некоторые языки, находящиеся под угрозой исчезновения, с давних времен пользуются письмом, в то время как другие являются только устными.

Некоторые из них считаются «семейными языками», которыми не принято делиться с посторонними. Для обозначения объектов, понятий и идей, необходимых для оцифровки языка, защитники языков коренных народов должны документировать существующие и создавать новые слова.

Целевая аудитория этого руководства — коренные народы, желающие документировать свой лексический запас, эксперты в разных предметных областях, проводящие терминологическую работу на одном или нескольких языках, и организации, поддерживающие языковые сообщества в процессе оцифровки. Конечная цель — получить возможность использовать собранную терминологию в устном, письменном и, в результате, цифровом формате.

Описанные рабочие процессы могут применяться к терминологической работе в любой тематике или предметной области. Приведенные здесь примеры в значительной степени касаются английского языка и будут применимы не во всех случаях, но некоторые общие правила и модели могут применяться к различным языкам и ситуациям. Эти примеры могут послужить хорошей отправной точкой для определения параметров, которые необходимо учитывать в конкретном языке при проведении терминологической работы.

В данном документе описываются основные области знаний, передовой опыт и ресурсы (как технологические, так и иные), которые помогут коренным народам

проводить терминологическую работу для сохранения, возрождения и оцифровки их языка.

### 3. 3. ЧТО ТАКОЕ ТЕРМИНОЛОГИЯ И ТЕРМИНОЛОГИЧЕСКАЯ РАБОТА?

Специальная терминология языка состоит из слов, аббревиатур и даже предложений, используемых в языке для специальных целей. Конечно, наиболее широко известна научная и техническая терминология, но во всех культурах все специализированные виды человеческой деятельности используют специальную терминологию, будь то приготовление пищи, охота или выражение культурных ценностей. Цели терминологической работы:

- Собираение слов и терминов: собрать и документировать существующие слова и термины. В случае коренных и недокументированных языков эта деятельность является частью языковой документации и языкового планирования, при которой собираются все возможные слова и термины. Этот процесс не ограничивается специальными языковыми терминами как таковыми; где это уместно, предметным областям также присваиваются понятия.
- Введение новых терминов для отсутствующих понятий: если возникает необходимость в понятиях, которые в языке еще не существуют, для создания новых терминов по возможности используются уже записанные, существующие ресурсы языка. Терминологи обычно находят такие понятия, сравнивая похожие понятия и обнаруживая, что понятие, которое они знают в одном языке, не существует или не установилось в другом. Правильное обозначение таких пустых слотов — «отсутствующие термины». Иногда эти слоты заполняются путем заимствования. В английском языке есть такие слова, как *tsunami* и *Schadenfreude*, заимствованные из других языков. В

случае развивающихся языков обычно рекомендуется не заимствовать, а создавать новые термины, используя существующие ресурсы.

Работа с терминологией играет важную роль в сохранении культуры и продвижении равноправия. Когда языки находятся под угрозой исчезновения, терминологическая работа помогает документировать и возродить язык. Например, валлийский язык или современный иврит находились на грани исчезновения, но стали основными национальными языками, отчасти благодаря систематической работе с терминологией.

Терминологическая работа проводится для того, чтобы собирать слова, которые действительно в обиходе, и фиксировать общее понимание их значений. Собрание таких слов из разных областей специализированной деятельности представляет собой основу, которая дает носителям языков коренных народов возможность активно участвовать в процессе документирования своего языка, сохраняя точность терминологии и ее надлежащее применение. Как отмечалось выше, при внедрении новых технологий работа над терминологией включает в себя создание новых терминов, что часто необходимо в случае языков коренных народов, которые преобразуются в цифровую форму или готовятся к использованию на цифровых устройствах. Такие термины цифрового пространства, как «браузер» или «инструмент», могут быть заимствованы из повседневных слов, но такие термины, как «меню», «шрифт» и даже «электронная почта», заставят задуматься, поскольку их корневые значения могут быть незнакомы.

#### 4. 4. ИДЕАЛЬНЫЙ РАБОЧИЙ ПРОЦЕСС

Ниже описан идеальный рабочий процесс терминологической работы, связанной с цифровизацией языков. Все этапы важны для выполнения этой задачи, но следовать им в указанном порядке может оказаться невозможным. Этот список служит приблизительным руководством для вашей работы.



#### 4.1 Вы готовы начать работу над терминологией?

- Имеет ли ваш язык все необходимое для цифровизации? Например:
- Есть ли в вашем языке система письма?
- Доступна ли ваша система письма в [стандарте Unicode](#) для записи терминологии с помощью компьютерных приложений?
- Настроена ли ваша система письма для ввода и отображения данных с клавиатуры на компьютерах и цифровых устройствах?
- Для вашего языка уже разработаны основные правила грамматики?

Если ответ на любой из этих вопросов «НЕТ», перейдите на вкладку Translation Commons Resources (Ресурсы Translation Commons) для получения дополнительной информации по решению этих проблем. Не останавливайтесь на достигнутом. Даже если некоторые из этих шагов уже выполняются, вы тем временем можете заняться другими вопросами. Информация в этом руководстве может повторяться, но она сосредоточена на процессе определения, записи и создании новых терминов. Имейте в виду, что вы можете начать делать аудиозаписи задолго до окончательного оформления письменности или системы письма. При этом начать использовать разрабатываемую систему письма для записи информации вручную можно еще до завершения процесса цифровизации.

#### 4.2 Вы знаете, кто будет выполнять эту работу?

- Кто из представителей вашего сообщества готов работать над проектом?
- Заручились ли вы поддержкой старейшин и других вождей?
- Кто из носителей вашего языка знаком с вашими темами? Есть ли в сообществе уважаемый повар, охотник или другой эксперт? Свяжитесь с ними и предложите им рассказать вам о своей работе.

- Знаете ли вы технических и компьютерных специалистов, которые могут оказать поддержку проекту?
- Знаете ли вы и ваша команда, как распознавать и собирать слова и термины?
- Чтобы собрать команду и объяснить всем ваши цели, может потребоваться некоторое время, особенно если вы ищете экспертов из разных предметных областей. Это непрерывный процесс, который может занять длительное время. Используйте обучающие и другие материалы, чтобы научить других распознавать термины и записывать их значения. Если вы, например, разрабатываете грамматику и стандарты орфографии, чтобы научить других применять основные грамматические правила, добавьте эту информацию в свою программу обучения.

Во время работы обязательно консультируйтесь с членами вашего сообщества. Согласны ли они с вашими решениями? Есть ли разница в произношении или способах написания слов в вашей системе письма? В процессе работы вы можете столкнуться с диалектами. Все эти вопросы нужно рассмотреть, чтобы достичь согласия в вашей группе.

#### 4.3 Создавайте тексты и собирайте существующие термины на вашем языке.

- Лучше всего работать над одной конкретной темой, но будьте готовы документировать все новые слова и понятия, с которыми вы сталкиваетесь, указывая предметную область по мере работы.
- Узнайте, существуют ли уже какие-либо глоссарии, словари или тексты на вашем языке, относящиеся к теме, с которых вы можете начать сбор терминов.
- Поговорите с wybranymi экспертами, сделайте аудиозаписи их историй и комментариев.

- Записывайте от руки, если это необходимо. Позже расшифруйте свои записи.
- Выделяйте важные слова и термины. После этого начните составлять глоссарии и записывать определения или пояснения.
- Если при документировании терминологии вашего региона вам попадаются предметы, термины для которых вам неизвестны, попробуйте найти их и включить в свои списки.
- Насколько точны ваши термины? Уточните у носителей языка, нет ли у терминов других форм. Существуют ли диалекты языка?
- Фиксируйте соответствующие термины по ходу работы.
- Храните свои тексты и глоссарии в одном месте, чтобы у вас был корпус текстов.
- Когда у вас будет достаточно большой корпус, возможно, вы захотите воспользоваться специальными компьютерными инструментами для организации корпуса и документирования терминологии.

#### 4.4. Придумывайте термины для описания новых идей и понятий.

- В вашем языке есть понятия, для которых еще нет терминов?
- Организуйте соответствующие термины в системы понятий, с тем чтобы можно было изучить взаимосвязь понятий, определить недостающие термины и сформулировать их определения.
- Нужно ли вам создавать целые группы терминов, например, по медицинской или компьютерной тематике?
- Спросите у нескольких носителей языка, понимают ли они придуманные вами термины.

#### 4.5 Фиксируйте и актуализируйте терминологию в общих глоссариях.

- Вы уже решили, как будете отслеживать термины и их значения?
- Вы определились с типами данных, которые будете документировать по каждому термину?

- Вы определились с системой управления терминологией и моделью данных, которые будете использовать для документирования терминов?
- Зафиксировано ли каждое важное понятие в сборнике терминов, структурированном в виде терминологической базы?
- Проводилась ли проверка качества каждой записи понятия?
- Предоставлялась ли терминологическая база пользователям, которым она нужна?
- Приводятся ли в вашей терминологической базе результаты ваших исследований?
- Можно ли из вашей терминологической базы понять, как использовать термины? Это называется прескриптивной терминологией.
- Обеспечивается ли единство понимания понятий из терминологической базы представителями соответствующих языковых сообществ?
- Нужно ли актуализировать терминологию с течением времени?

## 5. ПРИМЕНЕНИЕ УКАЗАНИЙ РУКОВОДСТВА

### 5.1 Как пользоваться блок-схемой

Ответы на данные вопросы собраны в блок-схеме на рисунке 1. Имейте в виду, что это всего лишь упрощенное представление. На самом деле выполнять различные части проекта можно и в другом порядке или даже параллельно. Все проекты разные. Во время работы следует проявлять гибкость и творческий подход.

Обратите внимание, что второй пункт в правом столбце в основном касается буквенной письменности, когда символы представляют собой звуки языка без привязки к значению. Как вариант, можно работать с логографическим письмом, в котором символы языка представляют некое понятие или идею без привязки к произношению соответствующего слова или термина. Английский, французский и

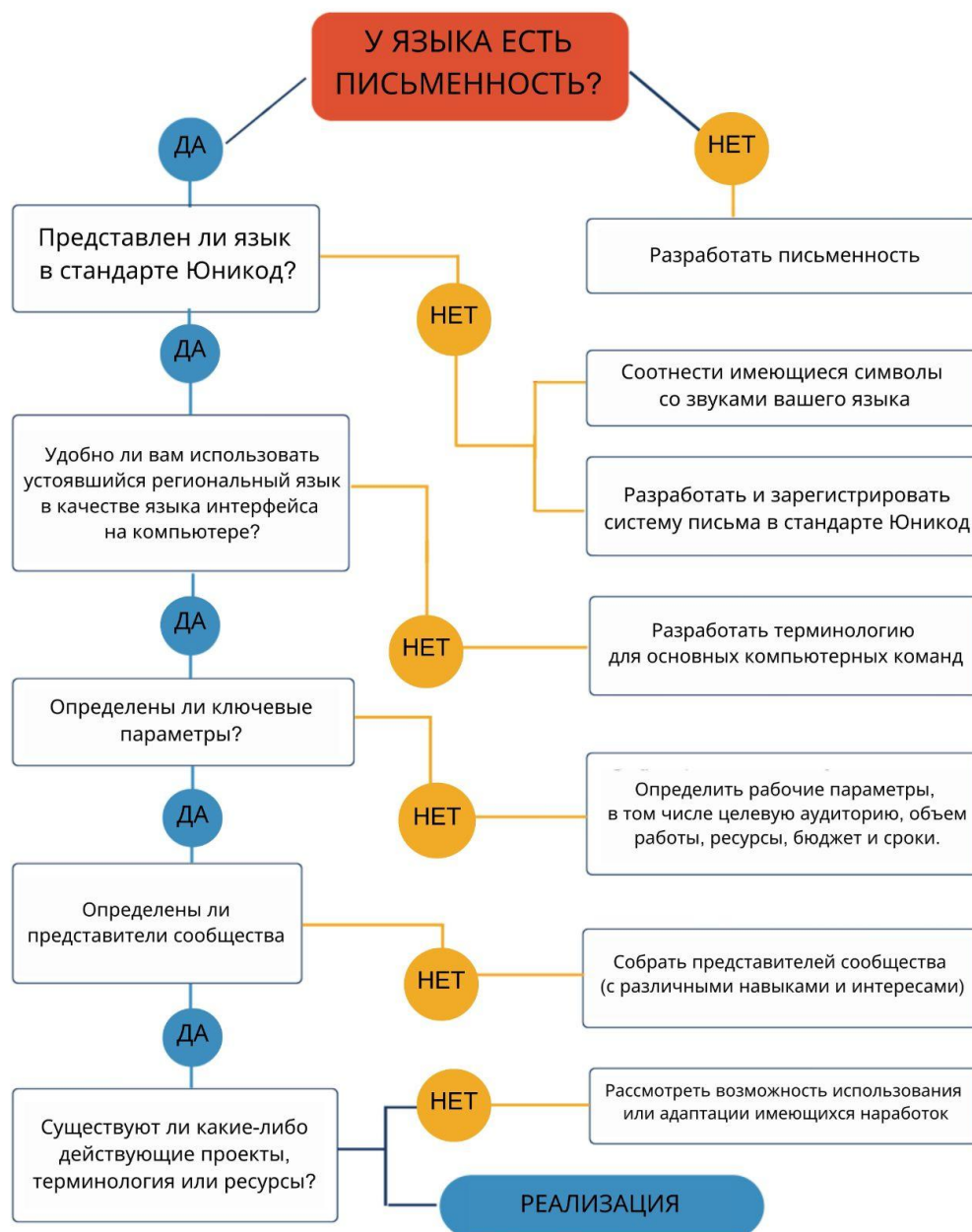
арабский языки — это примеры буквенной письменности, тогда как китайские и японские иероглифы представляют собой логографическое письмо.

Сначала, безусловно, нужно набрать людей, и когда вы определитесь с участниками, можно будет сразу приступить к работе с ними. По мере общения с людьми, формирования их доверия и развития сотрудничества вы можете начать делать аудиозаписи задолго до окончательного оформления письменности или системы письма. При этом начать использовать разрабатываемую систему письма для записи информации вручную можно еще до завершения процесса цифровизации. По мере того, как все большее количество людей будет общаться посредством мобильных телефонов и других цифровых инструментов, вы сможете использовать электронные ресурсы для сбора информации от более широкой группы людей, знающих язык.

Все это позволит сэкономить драгоценное время и получить исходную информацию, которую можно будет использовать для создания моделей предстоящей работы по мере развития проекта. Кроме того, такая предварительная работа обеспечит достоверные примеры текста, которые понадобятся позднее при подаче заявки на статус языка ISO 639, регистрации письменности или сертификации по стандарту Юникод. Сверяйтесь с блок-схемой по мере реализации проекта, чтобы не пропустить то, что не было сделано ранее.

# БЛОК-СХЕМА

## ПОДГОТОВИТЕЛЬНЫХ МЕРОПРИЯТИЙ



TRANSLATION  
COMMONS



Рис. 1. Блок-схема подготовки проекта

## 5.2 Лексикография (составление словарей известных понятий) и терминология (терминография)

Традиционные словари представляют собой списки слов в виде статей, когда в рамках одной статьи приводятся все значения (определения), связанные с этим словом (см. рисунок 2).

Лексикографические словари весьма полезны при работе с одним языком, поскольку они позволяют собрать всю информацию об отдельных словах в одном месте. В них также может содержаться определенная грамматическая информация и в них фиксируются многие короткие слова (так называемые пропускаемые слова, которые не учитываются при поиске в корпусах), например артикли, предлоги, союзы и другие языковые единицы, которые используются как в повседневном, так и в специализированном языке.

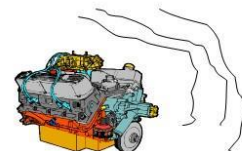
**rattle** (rat'l) vi. -tled, -tling, [ME ...]

1. to make a series of sharp, short sounds in quick succession
  2. to go or move with such sounds [a wagon rattling over the stones]
  3. to talk rapidly and incessantly; chatter [often with on: rattle on]
- vt.

1. to cause to rattle [to rattle the handle of a door]
  2. to utter or perform rapidly
  3. to confuse or upset; disconcert [to rattle a speaker with catcalls]
- n.

1. quick succession of sharp, short sounds
2. a rattling noise made by air passing through the mucous of a partially closed throat: cf. DEATH RATTLE
3. a noisy uproar; load chatter; engine rattle
4. a series of horny rings at the end of a rattlesnake's tail, used to produce a rattling sound
5. a device, as a baby's toy or a percussion instrument, made to rattle when shaken

Collocation: to rattle around in a house that is too big for one's needs



**Рис. 2. Словарная (лексикографическая) статья для слова rattle**

Однако лексикографические статьи довольно сложно использовать в качестве основы для двуязычных или многоязычных словарей или при документировании множества диалектных слов или синонимов. При поиске эквивалентов может оказаться так, что каждое значение словарной статьи в другом языке выражается совсем другим словом, а различные варианты написания или синонимы могут быть разбросаны по нескольким разным статьям. Поэтому на сегодняшний день большинство переводчиков, двуязычных авторов, исследователей и составителей словарей применяют так называемый терминологический подход к фиксации лексической информации.

### 5.3 Как определяется терминология, и почему эта работа так важна?

При использовании терминологического подхода в процессе документирования языковых норм мы создаем отдельную запись для каждого значения слова.

Эта задача может показаться сложной, особенно если посмотреть на рисунок 2, где при создании терминологических записей одна статья превращается сразу в одиннадцать. Но если это сделать, то каждому из значений можно будет легко присвоить термин, его синонимы и эквиваленты для нескольких языков или диалектов, то есть точно присвоить каждому понятию соответствующие эквиваленты. Такой подход нужен не только для точного перевода, но и для ориентира при создании новых терминов, которые необходимы для цифровизации или для формулирования различных отсутствующих понятий, например, в области медицины.

Для наглядного представления связи между тем или иным понятием и термином, которым обозначается это понятие, зачастую используют треугольник, углы которого показывают (1) понятие, с которым ассоциируется термин, (2) предмет, обозначаемый термином (так называемый референт, или предмет из реального мира), и (3) сам термин. Если термин выполняет свою функцию правильно, то мы можем использовать его в устной или письменной речи так, чтобы у слушателя



или читателя он ассоциировался с соответствующим понятием. Когда мы общаемся более чем на одном языке, важно, чтобы предмет (референт) в обоих языках был одним и тем же или очень близким по значению, чтобы все участники одинаково понимали одно и то же понятие, которое мы имеем в виду. Только в этом случае можно будет говорить, что термин на языке А является адекватным эквивалентом термина на языке В. Пример семантического треугольника, используемого для представления значения, показан на рисунке 3.

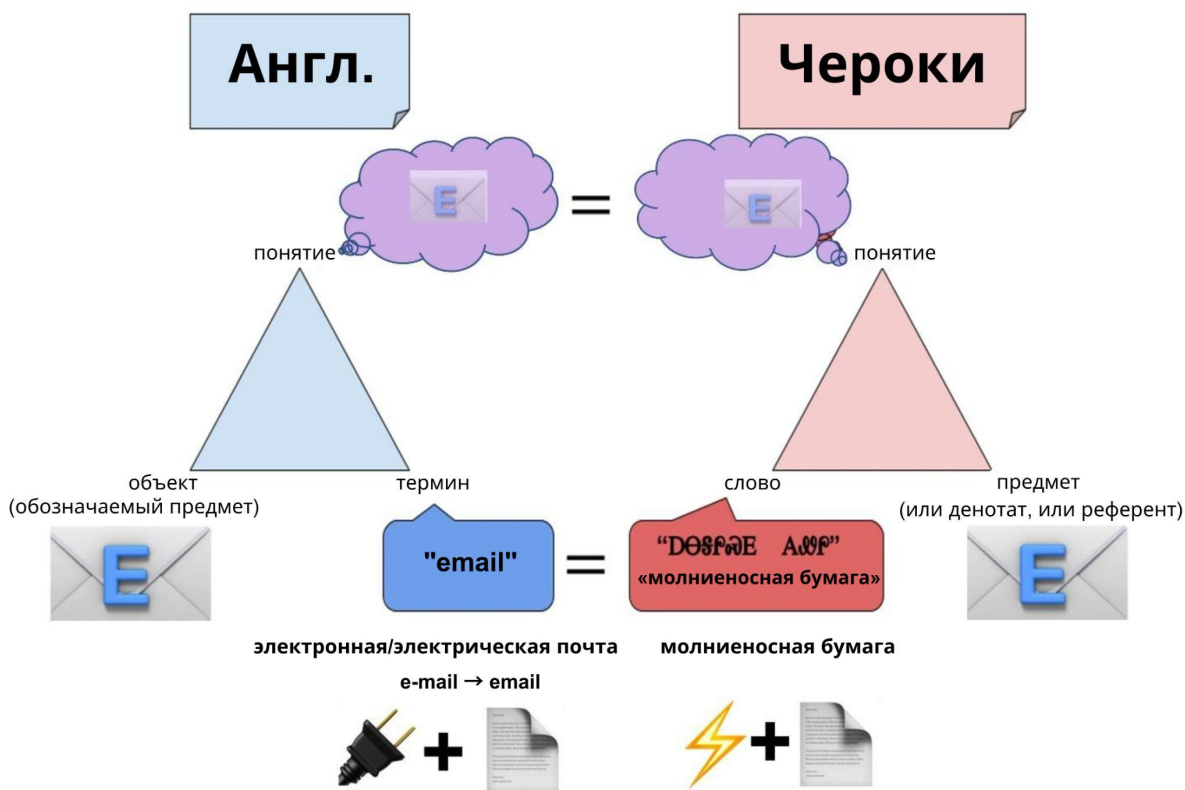
Такая схема треугольника особенно полезна при сравнении языков. Сначала мы уславливаемся, что слово представляет то или иное понятие. Затем мы ищем эквивалентное понятие в целевом языке. Если для этого понятия нет эквивалентного слова, возникает типичная ситуация «отсутствующего» слова. В этом случае можно создать новое слово или заимствовать его. В идеале терминологическая работа направлена на выявление существующих в языковом сообществе терминов, а не на заимствование слов из других языков. Удачным примером этому является термин tree («дерево»). В языке коренных американцев чероки есть слово, обозначающее дерево: ᏊᏂᏃᏍᏉ (quigvi [kʷigǽi]). Однако существует множество других английских слов, для которых в языке чероки изначально не было эквивалентов. Предположим, нам нужно придумать подходящий термин для понятия, обозначаемого английским словом email («электронная почта»). Если мы хотим избежать заимствования, мы не будем просто записывать слово на языке чероки, которое звучит как E-MAIL.



**Рис. 3. Семантический треугольник, в котором термин tree соответствует обозначаемому предмету (референту) — дереву, ассоциирующемуся с понятием дерева в сознании человека, который слышит или читает этот термин.**

В рамках терминологической работы мы исходим из того, что независимо от того, был ли у носителей английского языка и языка чероки одинаковый опыт работы с электронными письмами на момент создания терминов, само по себе электронное письмо обозначает один и тот же предмет для них обоих. Обозначаемый предмет (или референт) — электронное письмо — остается неизменным в этом уравнении. Иногда известные предметы, такие как деревья, могут выглядеть для двух говорящих немного по-разному (например, дуб выглядит иначе, чем баобаб), но когда мы называем или определяем наш предмет, мы смотрим на общие характеристики двух предметов. В первую очередь именно терминологический узел, скорее всего, будет отличаться и, вероятно, будет отражать различный опыт говорящих, читающих и пишущих. Сравните английское слово email, сокращенное от electronic mail («электронное письмо»), и слово на языке чероки DƏSʔƏE AʔP (anagalisgv goweli [anagalisgẽ goʔeli]), которое дословно переводится как молниеносная бумага. Хотя носителям английского языка и доводилось видеть

молнию, они изо дня в день имеют дело с электричеством, проходящим через электрическую систему (показана символом в виде электрической вилки), и они знают, что электронные письма доходят до них именно благодаря электричеству. Конечно, современные носители языка чероки тоже это понимают, но их язык не эволюционировал теми же темпами, что английский. В данном случае наблюдается конвергенция двух понятий: электричества, которое проходит по проводам, и электричества в виде молнии, а также скорости доставки и волшебных свойств электронной почты. Поэтому чероки решили использовать естественное для них слово для обозначения электричества в этом термине. Так в итоге и получилась «молниеносная бумага».



**Рис. 4. Треугольник, показывающий связь между понятиями в двух языках, когда совпадающие понятия выражаются эквивалентными терминами для одного и того же обозначаемого предмета (референта).**

Данный подход к сопоставлению термина с понятием весьма важен, поскольку, как уже отмечалось ранее, некоторые термины имеют более одного значения, а некоторые понятия могут быть представлены более чем одним термином, и простое сопоставление терминов друг с другом без привязки к понятию может оказаться ненадежным. Такой подход упрощает сопоставление понятий в двух и более языках, позволяет учитывать различные формы в нескольких диалектах и дает информацию о различных вариантах написания.

Основная причина актуальности двуязычных терминологических статей заключается в том, что термины из языков коренных народов зачастую документируются вместе с их эквивалентами на втором языке, будь то чероки и английский. Обычно мы избегаем заимствования, но если какие-то термины уже были заимствованы и вошли в широкое употребление, можно оставить их в списке терминов, а потом добавить в терминологическую базу. Однако по возможности лучше создавать термины, используя знакомые слова из собственного языка, а не заимствования. Такой подход помогает носителям языка понимать и управлять процессом цифровизации своего языка. Кроме того, преимущества подхода проявляются и в других областях, поскольку он позволяет создавать новые сущности. Например, использование знакомых слов в медицине помогает медработникам оперативно получать информацию, принимать необходимые меры и поддерживать контакт, что способствует достижению положительных результатов.

## 6. ДОКУМЕНТИРОВАНИЕ ТЕРМИНОВ

Списки терминов весьма полезны в начале работы, но еще вам понадобятся определения для ваших терминов (иногда даже по несколько определений на термин), и по мере роста ваших списков вам будет все труднее находить в них нужную информацию. Рано или поздно вы придете к необходимости создавать терминологические записи, особенно если вам нужно будет документировать термины с эквивалентами на других языках. Простая полная терминологическая

запись (модель данных для записи термина) может выглядеть примерно как на рисунке 5.

Как правило, в терминологическую запись включаются следующие элементы данных, обычно называемые категориями данных:

- **Предметная область**
- **Определение** понятия и его источник (при наличии)
- **Термин**
- **Часть речи**
- **Пример** типичного употребления термина в контексте (с указанием источника)

<b>Предметная область:</b>	цифровые коммуникации
<b>Термин на языке чероки:</b>	DƏSPƏDE AƏP
<b>Часть речи:</b>	существительное
<b>Определение:</b>	DƏSPƏDU AƏP: əZPədu DəduVəəduWə DƏSPədu DəYəduJədu EWəY ədu Hə EJədu DSVIəədu DəduThAJədu hLGŁədu ƏT Hə Də O'hAJ JhAJədu ƏtT TJP O'SLAP
<b>Контекст:</b>	Источник: <a href="https://language.cherokee.org">https://language.cherokee.org</a> Ə DFGG DƏSPədu AƏP əSəib OPT əZPədu ƆƏPəduLhRV. Источник: Roy Boney <a href="mailto:roy-boney@cherokee.org">roy-boney@cherokee.org</a>
<b>Термин на английском языке:</b>	Email
<b>Часть речи:</b>	существительное
<b>Определение:</b>	сообщение, передаваемое электронными средствами связи пользователем одному или нескольким получателям через компьютерную сеть Источник: <a href="https://www.google.com/search?q=define+email&amp;rlz=1C1CHBF_enUS866US866&amp;oq=define+email&amp;aqs=chrome..69i57j0i512i9.2230j0j15&amp;sourceid=chrome&amp;ie=UTF-8">https://www.google.com/search?q=define+email&amp;rlz=1C1CHBF_enUS866US866&amp;oq=define+email&amp;aqs=chrome..69i57j0i512i9.2230j0j15&amp;sourceid=chrome&amp;ie=UTF-8</a>
<b>Контекст:</b>	Я могу посмотреть свою электронную почту через телефон или с компьютера. Источник: Translation Commons

**Рис. 5.** Типичная схема терминологической записи

Для начала можно записывать всю эту информацию в редактируемый текстовый файл, но лучше использовать какую-нибудь программу для работы с электронными таблицами, позволяющую хранить термины в виде записей. Файл с электронной таблицей может выглядеть примерно так, как показано на рисунке 6. Два изображения на этом рисунке на самом деле представляют собой одну длинную строку в электронной таблице. Можно создавать электронные таблицы и с дроблением на более узкие сегменты, но если размещать каждое понятие на обоих (или даже всех) ваших языках в одну строку, потом будет заметно проще экспортировать данные в другие системы, если вы решите создать базу данных. Электронные таблицы хороши, когда нужно собрать небольшое количество терминов, но по мере своего роста они становятся громоздкими и непрактичными. Создав электронную таблицу, подходящую для модели терминологической базы, вы сможете относительно легко экспортировать и импортировать свои данные.

	A	B	C	D	E	F
1	<b>Cherokee term</b> Də́sɪnə́e A.ʋə́ŋ	<b>PoS</b> noun	<b>chr-Definition</b> Də́sɪnə́e A.ʋə́ŋ: ʋə́ŋə́e D.ə́ʋə́ŋə́e W.ə́ŋə́e D.ɪ́ʋə́ŋə́e E.W.ə́ŋə́e ʋə́ŋə́e E.ɪ́ʋə́ŋə́e D.S.V.ɪ́ʋə́ŋə́e D.ə́ŋə́e h.L.C.ɪ́ʋə́ŋə́e ʋə́ŋə́e D.ə́ŋə́e J.h.ɪ́ʋə́ŋə́e ʋə́ŋə́e T.J.P. ʋə́ŋə́e	<b>chr-Def-Source</b> <a href="https://language.cherokee.org">https://language.cherokee.org</a>	<b>chr-Context</b> ʋə́ŋə́e D.ə́ŋə́e A.ʋə́ŋə́e ʋə́ŋə́e ʋə́ŋə́e ʋə́ŋə́e ʋə́ŋə́e	<b>chr-Ctx-Source</b> Roy Boney <roy-boney@cherokee.org>
2						
3						

	G	H	I	J	K	L
1	<b>English term</b> email	<b>PoS</b> noun	<b>en-Definition</b> message distributed by electronic means from one computer user to one or more recipients via a network	<b>en-Def-Source</b> <a href="https://www.google.com/search?q=what+is+an+email&amp;rlz=1C1CHBF_enUS866US866&amp;oq=what+is+an+email&amp;aqs=chrome..69l57j0l512l9_3696j0j15&amp;sourceid=chrome&amp;ie=UTF-8">https://www.google.com/search?q=what+is+an+email&amp;rlz=1C1CHBF_enUS866US866&amp;oq=what+is+an+email&amp;aqs=chrome..69l57j0l512l9_3696j0j15&amp;sourceid=chrome&amp;ie=UTF-8</a>	<b>en-Context</b> I can access my email on my phone or from my computer.	<b>en-Ctx-Source</b> Translation Commons
2						

Рис. 6. Терминологическая запись в виде одной строки в программе для работы с электронными таблицами.

**Предметная область.** Данная категория в записи помогает упорядочить термины по темам. Обязательно придерживайтесь какого-то одного значения. Например, используйте только значение «Медицинская (Medical)», но не пишите произвольно то «Медицинская (Medical)», то «Медицина (Medicine)».

Придерживаясь одного подхода, вы сможете легко отфильтровать медицинские термины в своем ресурсе.

**Определение.** Как правило, хорошие определения — предельно короткие, но максимально точные. Если термин обозначает какую-то вещь, в определении обычно указывается принадлежность к более общему понятию, но с определенными отличительными характеристиками (например: волк — это крупное плотоядное животное семейства собачьих). В некоторых других случаях можно указать, что данный термин является частью более общего понятия (например, лист — это часть растения, которая является его органом, поддерживающим растение за счет фотосинтеза). Глаголы часто определяются пояснительной глагольной фразой. Например, ходить: перемещаться в обычном темпе, поднимая каждую ногу по очереди. В разделах 7.3 и 7.4 обсуждается создание схем системы понятий. Собрав достаточное количество терминов, вы сможете наглядно представить взаимосвязь между ними посредством таких схем. Для этого может потребоваться уточнение ваших определений, чтобы они отражали взаимосвязь между понятиями.

**Термин.** В начале работы над проектом вы можете просто записывать термины по мере их появления. В процессе изучения грамматики постарайтесь узнать, есть ли для ваших терминов что-то вроде начальной формы. Во многих языках чаще всего это существительное в единственном числе. При этом форма множественного числа может отличаться, а в некоторых языках разные формы еще и по-разному проявляют себя в предложении. Если форм много, следует выбрать наиболее оптимальный способ их записи на вашем языке. Простую начальную форму часто называют леммой. Примеры лемм в английском языке: run (в отличие от runs или running), book (в отличие от books или booked), red (в отличие от redden или redness, которые, кстати, являются разными понятиями).

**Часть речи.** В предыдущем абзаце речь шла только о существительных и глаголах, но еще есть небольшие служебные слова, например предлоги, союзы и прилагательные, которые обычно служат для описания существительных, а также

наречия, которые служат для описания глаголов, слов и фраз, не относящихся к существительным. Эти категории называются частями речи. Данная система может отличаться в зависимости от языка, поэтому в рамках анализа грамматики вашего языка придется учитывать все эти нюансы. Часть речи термина и другие грамматические характеристики, вероятно, лучше всего записывать в одноязычный словарь, а сами термины, независимо от их части речи, фиксировать в терминологической базе. При этом уникальные слова, особенно прилагательные и наречия, можно фиксировать в вашем терминологическом ресурсе.

Следует отметить, что иногда существительные в одном языке могут быть глаголами в другом и наоборот. В английском языке отглагольные существительные часто преобразуются в глаголы действия, а в других языках (например, потаватоми, язык американских индейцев) существительные, которые всегда находятся в состоянии движения, могут трактоваться как глаголы (например, слово, обозначающее реку). Возможно, вам придется потрудиться, чтобы продумать, как части речи работают в вашем языке и как лучше изменить способ записи информации в терминологической базе.

**Контекст.** В этом поле можно записать какой-нибудь удачный пример употребления термина в предложении. Для этого лучше всего подойдет контекст, прямо или косвенно указывающий на определение понятия, а также какие-нибудь распространенные идиоматические выражения или обороты, в которых используется этот термин. Возьмем для примера термин «книга». В предложении «Джон покупает книги всякий раз, когда приезжает в город» ничего не говорится о значении слова «книга», равно как и в предложении «Они уселись вместе и по очереди читали вслух рассказы о рыбаках Южно-Яванского моря».

**Примечание.** В образце записи нет поля «Примечание», но вы можете добавить его, если у вас есть какая-либо дополнительная информация о записи.



## 7. ПРОЧИЕ РАБОТЫ

### 7.1 Создание корпуса текстов

Существуют ли для вашего языка готовые корпуса текстов? У некоторых языков есть древние варианты письменности, которые только сейчас проходят процесс цифровизации. Это позволит сохранить письменность и сделать ее более доступной. Другие языки богаты устным народным творчеством: преданиями, сказаниями, музыкой и поэзией, которые веками передаются из уст в уста, но не зафиксированы в письменном виде и которые могут оказаться под угрозой исчезновения, если число носителей языка будет сокращаться, а молодые поколения перестанут запоминать тексты. Как отмечалось ранее, можно начать с аудиозаписи таких «текстов», а затем расшифровать их для цифровизации. В любом случае, записывая и собирая информацию в устной и письменной форме, вы постепенно наберете текстовый корпус. Вся ваша коллекция в целом может как раз и стать тем самым текстовым корпусом. Информацию о сборе корпусов, защите авторских прав и создании архивов можно найти среди [ресурсов](#) на сайте Translation Commons.

Существует множество компьютерных инструментов для работы с текстовыми корпусами, в том числе для извлечения терминов. Инструменты для извлечения терминов разработаны специально для выявления потенциальных терминов, возможных определений и контекстов при работе с основными языками. Собрав значительный корпус, вы сможете обратиться к программистам, с тем чтобы они настроили один или несколько таких инструментов для работы с вашим языком. При работе с текстами, представляющими культурное наследие вашей группы, всегда крайне важно беречь и уважать ее убеждения и традиции. Даже работая на такую благую цель, как сохранение священных или конфиденциальных материалов для будущих поколений, следует помнить о необходимости их защиты.

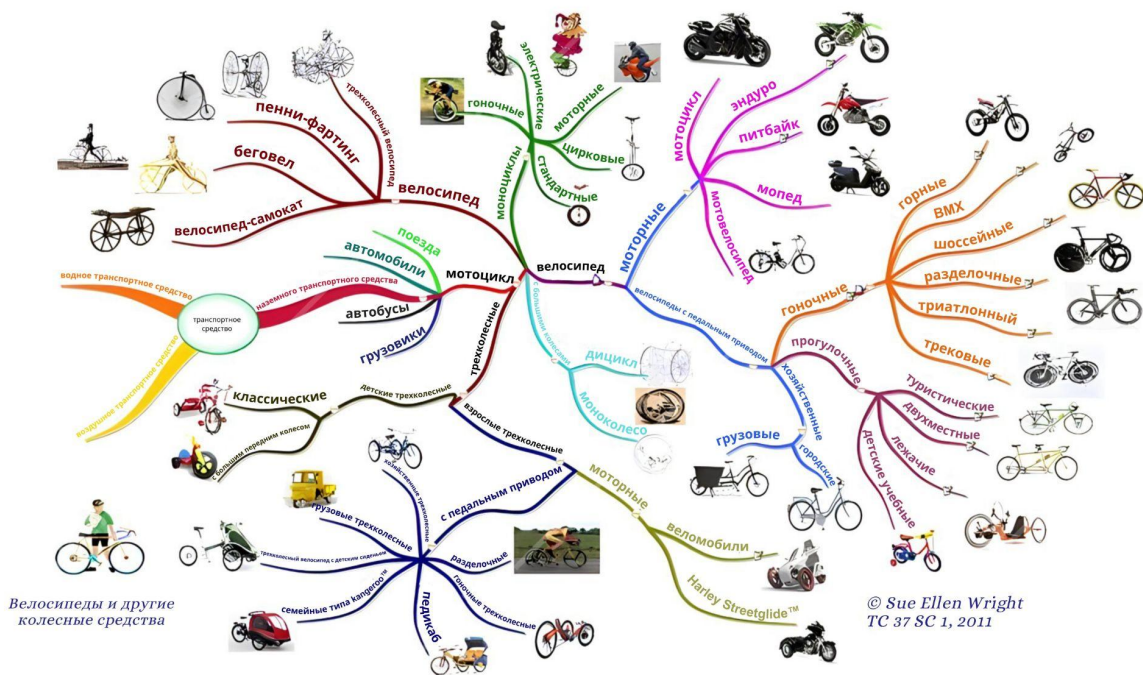
## 7.2 Добавление переводов для создания корпуса параллельных текстов

Ценность вашего корпуса сильно возрастет, если он будет выполнен на двух языках. Это позволит расширить ваши терминологические ресурсы как за счет человеческого, так и за счет машинного перевода (МТ). Такой двуязычный корпус создается путем перевода текста по отдельным предложениям с вашего языка на какой-либо другой язык. Тем самым обеспечивается строгое соответствие переведенных предложений и фрагментов предложений, что позволяет создать память переводов, состоящую из так называемых текстовых сегментов. Получившиеся таким образом тексты образуют корпус параллельных текстов. Для дополнения цифровых ресурсов машинным переводом сначала нужно собрать довольно большое количество переведенных документов, которые можно будет использовать для «обучения» вашей системы машинного перевода, чтобы она могла обнаруживать и извлекать последовательно переведенные термины и сегменты (см. «На пути в цифровое пространство. Руководство по машинному переводу»).

Первым важным шагом к внедрению решений для перевода является адаптация имеющихся инструментов автоматизированного перевода (CAT-инструментов) для обработки ваших текстов и правильного отображения вашего языка. Как только переводчики смогут использовать эти инструменты для создания памяти переводов (ТМ), вы сможете сохранять совпадающие сегменты, состоящие из предложений или фраз исходного языка, переведенных на другой язык. Эти сегменты можно использовать для сохранения и даже воссоздания целых документов, что позволит вам создать корпус параллельных текстов с соответствующим переводом. По мере увеличения вашего корпуса вы сможете обратиться к крупным поставщикам услуг машинного перевода, с тем чтобы они добавили ваш язык в список языков, поддерживаемых их системами. Кроме того, память переводов значительного размера можно использовать для извлечения не только терминов, но и большей части данных, необходимых для составления полных, содержательных терминологических статей.

### 7.3 Создание систем общих понятий

Специалисты по терминологии настоятельно рекомендуют в процессе работы систематизировать термины в системы понятий. Когда мы обсуждали определения, мы говорили об использовании конструкций типа «... — это» и «... — это часть». Задайте себе вопрос: что такое термин А? Если на него можно ответить, что термин А — это разновидность термина В, то можно смело приступать к созданию системы общих понятий. Например, многие знают, что такое велосипед. Если бы вам нужно было создать систему понятий, отражающую все возможные виды велосипедов, она могла бы выглядеть примерно как на рисунке 7. Такая система понятий — это отличный способ научиться определять понятия. Взглянув на рисунок, вы сможете начать формулировать определение, написав, что велосипед — это вид наземного транспортного средства с двумя колесами, которое обычно приводится в движение человеком с помощью педалей.



**Триатлонный велосипед — это вид гоночного велосипеда.**  
**оночный велосипед – это вид велосипеда**  
**Велосипед — это вид наземного транспортного средства.**

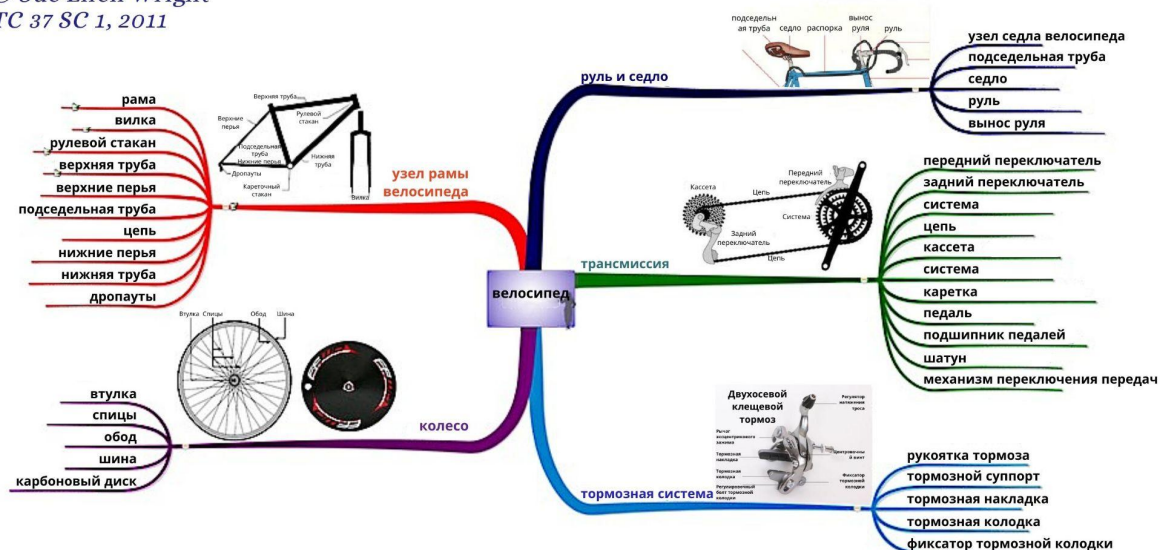
**Рис. 7. Системы общих понятий**

#### 7.4 Системы понятий «часть-целое»

Кроме того, понятия можно определить, исходя из того, как определяемый предмет делится на части. В примере на рисунке 9 педаль является частью трансмиссии велосипеда.

# Велосипедные детали и узлы

© Sue Ellen Wright  
TC 37 SC 1, 2011



**Колесная пара — это часть велосипеда.  
Колесо — это часть колесной пары.  
Спица — это часть колеса.**

Рис. 8. Системы понятий «часть-целое»

Построение системы понятий также позволяет выявлять отсутствующие термины в процессе сбора терминов или создания их эквивалентов. Зачастую системы понятий помогают обнаружить сразу несколько значений термина, о которых вы изначально даже и не думали.

## 8. СОВМЕСТНОЕ ИСПОЛЬЗОВАНИЕ И ПУБЛИКАЦИЯ

### ТЕРМИНОЛОГИЧЕСКИХ РЕСУРСОВ

Как отмечалось выше, чтобы увеличить коллекцию терминологии и сделать ее доступной для более широкой аудитории, лучше всего архивировать данные с помощью системы управления терминологией, работающей в режиме онлайн. Это может быть, например, программное обеспечение [Terminologue](#), которое было изначально разработано Fiontar &

Scoil na Gaeilge для [Foras na Gaeilge](#) в целях управления Национальной базой данных терминологии ирландского языка Téarma ([téarma.ie](#)). На сайте Terminologue указано, как настроить терминологическую базу и начать вносить данные.

Определите, кто из вашей команды будет отвечать за редактирование и администрирование терминологической базы. Сделайте терминологическую базу открытой для заинтересованных пользователей, чтобы они могли читать ее с компьютера, телефона и других цифровых устройств. Предоставьте гостям и заинтересованным пользователям средства для предложения новых записей и редактирования уже существующих, но не забудьте обеспечить защиту данных от нежелательного редактирования или непреднамеренных изменений.

## ИСПОЛЬЗОВАННЫЕ ИСТОЧНИКИ

### Пример проекта:

Patyegarang — сайт по обучению языкам коренных народов Австралии

<http://www.indigoz.com.au/language/gaps.html>

На этом сайте содержится подробная информация о создании новых терминов, основанная на опыте австралийского сообщества, которое долгое время занимается созданием терминов.

### Инструменты:

#### Инструменты для управления корпусами текстов

На сайте <https://www.Corpus-Analysis.com> представлен расширенный список из 261 инструмента для анализа корпусов.

<https://tesolpeter.wordpress.com/a-brief-guide-to-corpus-analysis-tools/>

<http://inmyownterms.com/readings-tools-and-useful-links-for-corpus-analysis/>

Существует множество инструментов и их набор постоянно обновляется. Многие из них доступны бесплатно, а на некоторые предоставляются специальные скидки для таких проектов.

#### Управление терминологией:

*Terminologue* — это инструмент для управления терминологией с открытым исходным кодом. Данное программное обеспечение разработано и поддерживается исследовательской группой [Gaois research group](#) in [Fiontar & Scoil na Gaeilge](#), Городской университет Дублина ([Dublin City University](#)). Программное обеспечение защищено авторским правом Дублинского городского университета и доступно по лицензии Массачусетского технологического института ([MIT](#)) с открытым исходным кодом. Ведущий разработчик — [Михал Болеслав Мечура](#). Скачать и установить программное [обеспечение](#) Terminologue вы можете с данного сайта.

<https://www.terminologue.org/docs/info.cs/>

## Инструменты для извлечения терминов:

### **Категории терминологических данных:**

Существует еще много категорий данных, которые можно использовать в терминологических записях. Найти их вместе с определениями и примерами использования вы можете на сайте <https://www.datcatinfo.net>.

### **Приложения для автоматизированного перевода (CAT-инструменты), как бесплатные, так и платные:**

<https://www.marstranslation.com/blog/top-free-and-paid-cat-tools>

Good Firms.

«10 бесплатных приложений для автоматизированного перевода с открытым исходным кодом»

<https://www.goodfirms.co/blog/the-top-10-free-and-open-source-computer-assisted-translation-software>

Инструмент для перевода с открытым исходным кодом Translate5

<https://www.translate5.net/en/translate5-open-source-translation-system-2>

Translation Commons

@TranslationCommons

@TranslationCom1





**ЛИНГВИСТЫ:**

Ksenia Yakovleva  
Stepan Konev  
Anastasija Artiomova  
Stanislav Bilan

**Менеджеры проектов по локализации:**

Chen Yao, Yutong Du,  
Sophie Liu, Han Liu,  
Liuyi Yang, Jiaqian Wang,  
Di Wan, Yiwei Wu

**Инженеры по локализации:**

Luiyi Yang  
Jiaqian Wang  
Yiwen Wang