

Серия “От нуля к цифре”

# Руководство по терминологии

TRANSLATION  
COMMONS



2022-2032 | МЕЖДУНАРОДНОЕ ДЕСЯТИЛЕТИЕ  
ЯЗЫКОВ КОРЕННЫХ НАРОДОВ

## Руководство по терминологии

Авторы: Сью Эллен Райт, Акил Айер, Шуто Като  
Рецензенты и соавторы: Крэйг Корнелиус, Алайна Брандт  
Джули Андерсон, Текс Тексин  
Маркетинг: Леонидас Паппас  
Координация проекта: Паула Сирило

Мы открыты для предложений по улучшению  
руководств.

Напишите нам письмо по электронной почте  
[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. ПРЕДИСЛОВИЕ	4
2. О ДАННОМ ДОКУМЕНТЕ	6
3. ЧТО ТАКОЕ ТЕРМИНОЛОГИЯ И ТЕРМИНОЛОГИЧЕСКАЯ РАБОТА?	7
4. ИДЕАЛЬНЫЙ РАБОЧИЙ ПРОЦЕСС	9
4.1 Вы готовы начать работу над терминологией?	10
4.2 Собрали ли вы людей для выполнения этой работы?	10
4.3 Создавайте тексты и собирайте существующие термины на вашем языке.	12
4.4. Придумывайте термины для отсутствующих новых идей и понятий.	12
4.5 Документируйте и обновляйте терминологию в общих глоссариях.	13
5. ПРИМЕНЕНИЕ РУКОВОДСТВА	14
5.1 Использование технологической карты	14
5.2 Лексикография (знакомые словари) и терминология (терминография)	17
5.3 Как мы определяем терминологию и почему так важна работа с терминологией?	18
6. ДОКУМЕНТИРОВАНИЕ ВАШИХ ТЕРМИНОВ	22
7. ДРУГАЯ ДЕЯТЕЛЬНОСТЬ	27
7.1 Создание текстового корпуса	27
7.2 Добавление переводов для создания параллельного корпуса	28
7.3 Создание систем общих понятий	29
7.4 Системы понятий «часть-целое»	30
8. СОВМЕСТНОЕ ИСПОЛЬЗОВАНИЕ И ПУБЛИКАЦИЯ ВАШИХ ТЕРМИНОЛОГИЧЕСКИХ РЕСУРСОВ	31
ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА	33

## 1. ПРЕДИСЛОВИЕ

[Translation Commons](#) — некоммерческое сообщество волонтеров, деятельность которого направлена на поддержку процесса перевода языков в цифровой формат, наставничество представителей языковых профессий, а также предоставление образовательных и информационных ресурсов в сфере языковых услуг.

Одной из центральных программ Translation Commons является «Инициатива по цифровизации языков», которая дает возможность заинтересованным языковым сообществам получить доступ к инструментам цифровых технологий. В мире существует около 6 000 языков с незначительным присутствием в информационном цифровом пространстве или вовсе в нем не представленных. В «Инициативе по цифровизации языков» предлагается стратегический план, следуя которому любое сообщество сможет осуществить цифровизацию своего языка.

Translation Commons выступали партнером [Международного года языков коренных народов мира 2019](#), инициативы ЮНЕСКО, для привлечения большего внимания к коренным народам и цифровизации их языков. Частью миссии «Инициативы по цифровизации языков» является обеспечение равных возможностей доступа к цифровым технологиям для носителей языков коренных народов и других языковых меньшинств с целью повышения сетевой активности таких сообществ и предоставления им возможности пользоваться современными компьютерными приложениями на своем родном языке. В созданных руководствах и информационных ресурсах по цифровизации шрифта и увеличению доли присутствия языков коренных народов в сети Интернет сообществам даются необходимые знания для самостоятельного преобразования языка в цифровой формат. В дополнение к руководствам, Translation Commons предлагает уроки и семинары, а также помогает сообществам в оцифровке языков, знакомя их с отраслевыми экспертами, которые помогают им с процессом стандартизации.

Данный документ входит в серию методических руководств под названием *«На пути в цифровое пространство»*, в котором комплексно рассматривается деятельность по цифровизации языков. Авторами этих руководств выступили специалисты по вычислительной технике и лингвисты. Этот документ предназначен для любого языкового сообщества, для которого возможное применение родного языка в цифровых системах представляет интерес.

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»](#). На вебсайте Translation Commons в подразделе проекта Language Digitization (Цифровизация языков) в разделе [Resources](#) (Ресурсы) размещены дополнительные материалы, посвященные «Инициативе по цифровизации языков» (руководства, презентации, видео и прочие документы).

## 2. О ДАННОМ ДОКУМЕНТЕ

В данном документе описан порядок записи терминологии на существующем, недокументированном или недостаточно документированном языке. В нем описаны процессы записи существующих слов и создания новой терминологии для поддержки общения на специфические темы, для которых в языках коренных народов может не существовать определенных слов. Эта методология носит рекомендательный характер и является руководством по различным рабочим процессам для создания или обновления терминологии, включая планирование, сбор ресурсов, терминологический анализ и документирование. Благодаря этому процессу коренные народы могут придумывать термины по мере необходимости, например, для технологических, медицинских, политических, юридических или

других областей, которые в противном случае могут быть вне исторической сферы их культуры или языка. Расширение и обновление такого лексического запаса поможет коренному народу получить доступ и использовать ресурсы, способствующие их включению в цифровую среду.

Некоторые языки, находящиеся под угрозой исчезновения, имеют давние традиции письменности, а другие являются строго устными. Некоторые из них рассматриваются как «семейные языки», которыми не принято делиться с посторонними. Защитникам языка необходимо документировать уже существующие слова и создавать новые слова для обозначения объектов, концепций и идей, которые понадобятся их народу, если они захотят оцифровать свой язык.

Целевой аудиторией этих рекомендаций являются коренные народы, желающие задокументировать свой лексический запас, эксперты в предметной области, поддерживающие терминологическую работу на одном или нескольких языках, и организации, поддерживающие языковые сообщества в процессе оцифровки. Цель заключается в получении возможности использовать полученную коллекцию терминов для устных, письменных и, в итоге, цифровых ресурсов.

Описанные рабочие процессы могут быть применены к терминологической работе в любой области или предметной области. Приведенные здесь примеры в значительной степени касаются английского языка, но некоторые общие правила и модели могут применяться к различным языкам и ситуациям, хотя некоторая информация будет применима не ко всем языкам. При этом, примеры должны дать читателям хорошую отправную точку для определения параметров, которые необходимо учитывать в их языке при проведении терминологической работы.

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

народам при проведении терминологической работы по сохранению, возрождению и оцифровке своего языка.

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

Специальные слова, аббревиатуры или даже фразы, используемые в языках, считаются частью специальной языковой терминологии данного языка. Конечно, терминология науки и техники широко известны, но все специализированные виды человеческой деятельности во всех культурах используют специальную терминологию, будь то приготовление пищи, охота или выражение культурных ценностей. Работа по терминологии ведется с двумя целями:

- Сбор слов и терминов: мы собираем и документируем существующие слова и термины. В случае коренных и недокументированных языков эта деятельность является частью языковой документации и языкового планирования, и мы собираем все слова и термины, какие только можем. В этой среде мы не ограничиваемся специальными языковыми терминами как таковыми, но присваиваем понятия предметным областям, где это уместно.
- Создание новых терминов для отсутствующих понятий: если это возможно, то мы используем существующие ресурсы языка, которые мы записали новых терминов, когда мы обнаруживаем, что есть понятия, о которых нам нужно говорить, которые еще не имеют терминов в языке. Терминологи часто определяют эти понятия, сравнивая похожие понятия и понимая, что понятие, которое они знают в одном языке, не было создано или придумано в другом. Правильное обозначение этих пустых слотов — «отсутствующие термины». Иногда эти места заполняются за счет заимствования. В английском языке есть такие слова, как *tsunami* и *Schadenfreude*, которые были заимствованы из других языков. При разработке языков обычно

считается разумным создавать новые термины, используя существующие ресурсы, а не заимствовать их.

Работа с терминологией важна для сохранения культуры и продвижения единообразия. Когда языки находятся под угрозой исчезновения, терминологическую работу можно использовать для создания записи о языке и возрождения его использования. Например, валлийский и современный иврит. Им грозило исчезновение, но они стали основными национальными языками, отчасти благодаря систематической работе с терминологией.

Терминологическая работа проводится для того, чтобы собрать слова, которые люди действительно используют, и зафиксировать общее понимание их значений. Набор таких слов из разных областей специализированной деятельности дает носителям языка коренных народов основу для активного участия в процессе документирования своего языка, сохраняя точность терминологии и их надлежащее применение. Как отмечалось выше, работа над терминологией включает в себя создание новых терминов при внедрении новой технологии, что характерно для многих языков коренных народов, которые оцифровываются или готовятся к использованию на цифровых устройствах. Такие цифровые термины, 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. ИДЕАЛЬНЫЙ РАБОЧИЙ ПРОЦЕСС

Ниже описан идеальный рабочий процесс терминологической работы, связанной с оцифровкой языков. Для завершения проекта важны все этапы, 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?

- Есть ли в вашем языке функции, необходимые для оцифровки? Например:
- Есть ли в вашем языке система письма?
- Доступна ли ваша система письма в [стандарте Unicode](#) для записи вашей 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](#) для получения дополнительной информации о решении этих проблем. Но не останавливайтесь на достигнутом. Даже если некоторые из этих шагов уже выполняются, вы можете заняться другими вещами. Информация в руководстве может повторяться, но она сосредоточена на процессе определения и записи терминов, а также на создании новых. Важно начать делать аудиозаписи еще до того, как письменность или система письма будут окончательно утверждены. Вы можете начать использовать развивающуюся систему письма для записи информации вручную до завершения процесса оцифровки.

#### 4.2 Собрали ли вы людей для выполнения этой работы?

- Собрали ли вы представителей вашего сообщества для работы над вашим проектом?
- Заручились ли вы поддержкой старейшин и других руководителей групп?
- Кто из носителей вашего языка знает о ваших темах? Кто у вас известный повар, охотник или другой эксперт? Свяжитесь с этим поваром или охотником и предложите им рассказать вам о своей работе.
- Вы нашли технических специалистов, которые могут предоставить поддержку проекту?

- Знаете ли вы и ваша команда, как распознавать и собирать слова и термины?
- Может потребоваться некоторое время, чтобы собрать команду, особенно если вы собираете экспертов в разных [предметных областях](#), и чтобы они смогли понять ваши цели. Этот процесс может занять долгое непрерывное время. Используйте учебные пособия и другие материалы, чтобы научить людей распознавать термины и записывать их значения. Если вы разрабатываете грамматику и стандартизируете орфографию, например, и людям нужно научиться применять основные правила грамматики, добавьте эту информацию в свою программу обучения.

Во время работы обязательно проверяйте свою работу с участниками вашего сообщества. Согласны ли они с вашими решениями? Есть ли разница в произношении или методах представления слов в вашей письменности? Вы можете обнаружить, что по мере того, как вы работаете, появляются диалекты. Нужно рассмотреть все эти вопросы, чтобы прийти к консенсусу 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 новое, что появляется на вашем пути, указывая предметное поле по мере работы.
- Определите, есть ли у вас уже какие-либо списки слов, словари или тексты на вашем языке, относящиеся к теме, которые вы можете использовать для начала сбора терминов.
- Поговорите с выбранными экспертами, сделайте аудиозаписи их историй и комментариев.
- Записывайте от руки, если это необходимо. Позже расшифруйте свои записи.

- Выделите важные слова и термины. После этого начните составлять списки и записывать [определения](#) или пояснения.
- Если вы документируете термины из своей гостинной и не знаете терминов для некоторых объектов перед глазами, попробуйте найти эти термины, чтобы включить их в свои списки.
- Ваши термины точны? Проверьте с несколькими носителями языка, есть ли другие формы? Существуют ли диалекты?
- Документируйте связанные термины по ходу работы.
- Храните ваши тексты и списки слов вместе, чтобы создать текстовый [корпус](#).
- Для больших корпусов существуют [инструменты](#) для их организации или документирования ваших терминов.

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

- Существуют ли нужные вам термины, которых нет в вашем языке?
- Организуйте связанные термины в системах понятий для изучения взаимосвязей понятий, определения недостающих терминов и продолжения написания [определений](#).
- Вам нужно создать отдельные группы терминов, например, для медицинских или компьютерных терминов?
- Посоветуйтесь с носителями языка, чтобы понять, значат ли что-нибудь придуманные термины.

#### 4.5 Документируйте и обновляйте 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?
- Задokumentировано ли каждое важное понятие в сборнике терминов в терминологической базе?

- Было ли проверено качество каждой записи понятия?
- Была ли терминологическая база предоставлена пользователям, которым она требуется?
- Указано ли в вашей терминологической базе то, что вы нашли в ходе исследования? Это называется *описательной терминологией*.
- Могут ли люди понять по вашей терминологической базе, как они должны использовать термины? Это называется *описательной терминологией*.
- Правильно ли соответствующие сообщества носителей ссылаются на понятия в терминологической базе?
- Нужно ли обновлять терминологию с течением времени?

## 5. ПРИМЕНЕНИЕ РУКОВОДСТВА

### 5.1 Использование технологической карты

Ответы на данные вопросы отображены на технологической карте на рисунке 1. Помните, что это примерный вид. На практике вы можете выполнять отдельные части проекта в разном порядке или параллельно. Все проекты разные. Во время работы стоит проявлять гибкость и творческий подход.

Обратите внимание, что второй пункт в правом столбце преимущественно для буквенной письменности, где символы представляют собой звуки языка без привязки к значению. Как вариант, модно работать с логографическим шрифтом, где символы языка представляют [понятие](#) или идею без привязки к тому, как произносится соответствующее слово или термин. Английский, французский и арабский языки — это примеры буквенной письменности. Китайский и японский кандзи — логографической.

Безусловно, в самом начале нужно собрать людей. После их выбора, вероятно, вы захотите начать с ними работу как можно скорее. По мере того, как вы разговариваете с людьми, завоевываете их доверие и развиваете сотрудничество,

вы можете начать делать аудиозаписи еще до того, как письменность или система письма будут окончательно утверждены. Вы можете начать использовать развивающуюся систему письма или письменность для записи информации вручную до завершения процесса оцифровки. Сейчас все больше людей общаются через мобильные телефоны и другие цифровые инструменты. Вы можете использовать электронные ресурсы для сбора информации от более широкой группы людей, знающих язык.

Все это сэкономит драгоценное время и предоставит исходную информацию, которую можно использовать для создания моделей будущих действий по мере развития проекта. Такая предварительная работа также даст достоверные примеры текста, которые понадобятся вам позже, когда вы подадите заявку на статус языка ISO 639, регистрацию письменности или сертификацию Unicode. По мере продвижения проекта проверяйте технологическую карту, чтобы понять, надо ли наверстать упущенное в предыдущих шагах.

# БЛОК-СХЕМА

## ПОДГОТОВИТЕЛЬНЫХ МЕРОПРИЯТИЙ

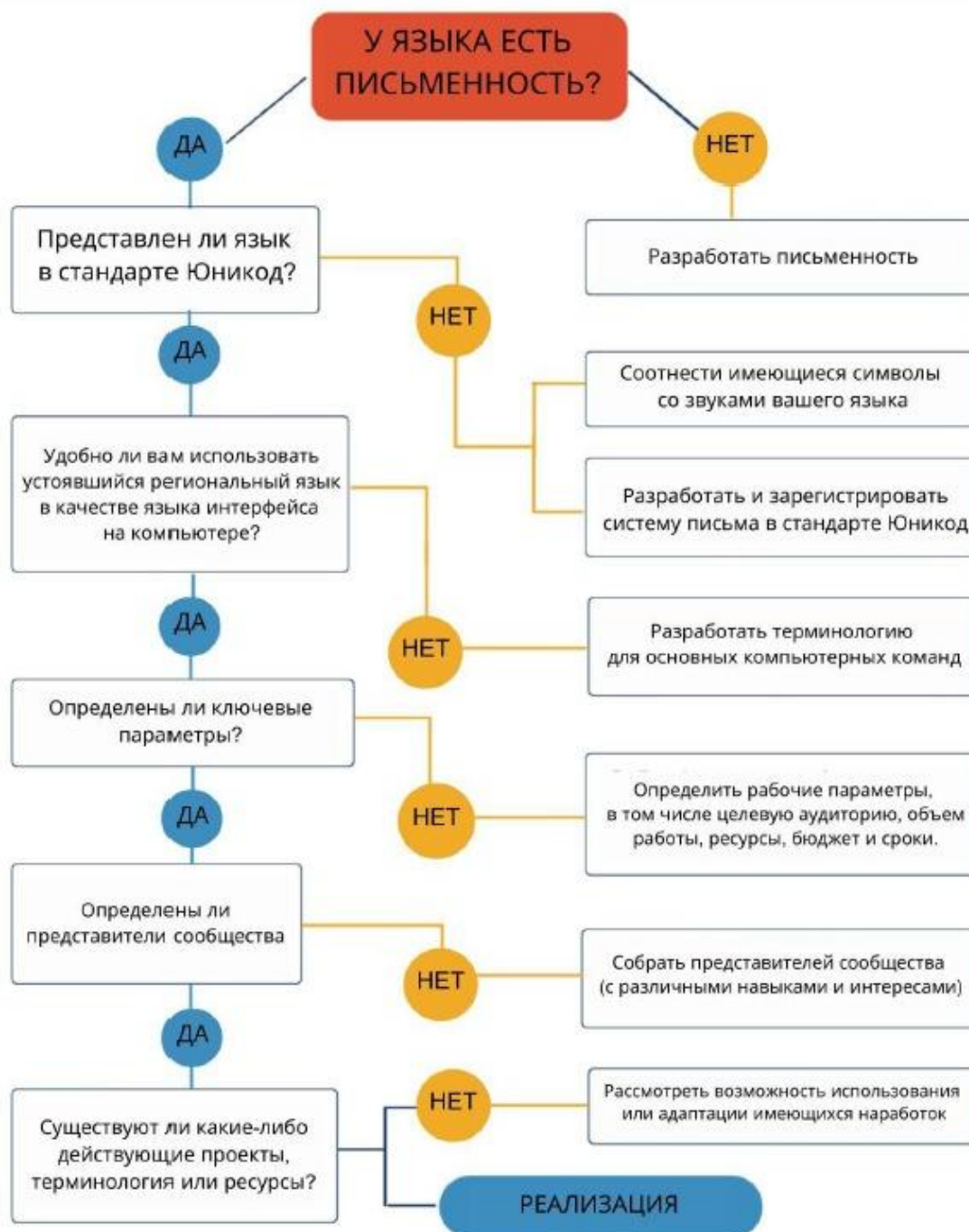


Рисунок 1: Технологическая карта подготовки проекта

## 5.2 Лексикография (знакомые словари) и терминология (терминография)

Традиционные словари перечисляют слова в статьях, где все значения ([определения](#)), связанные с этим словом, включены в одну и ту же статью, как показано на рисунке 2.

Лексикографические словари очень полезны, когда вы работаете только с одним языком, потому что вы можете собрать всю информацию об одном слове вместе в одном месте. Они также могут включать грамматическую информацию и используются для документирования многих маленьких слов (иногда называемых шумовыми словами или стоп-словами), таких как артикли, предлоги, союзы и другие частей речи, которые используются как в повседневном, так и в специализированном языке.

**rattle** (rat'ɪ) 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]

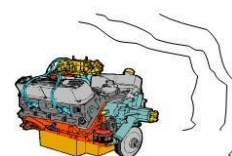
-vt.

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



### Рисунок 2: Словарная (лексикографическая) статья для слова «погремушка»

Однако лексикографические статьи становятся проблематичными, когда вы пытаетесь использовать их в качестве основы для двуязычных или многоязычных словарей или если вы хотите задокументировать множество

диалектных слов или синонимов. Каждое значение в статье потенциально может оказаться другим словом, когда вы ищете эквиваленты в другом языке, и разные варианты написания или в нескольких статьях могут оказаться синонимы. Поэтому сегодня большинство переводчиков, двуязычных авторов, исследователей, а также составителей словарей используют так называемый «терминологический подход» к записи лексической информации.

### 5.3 Как мы определяем терминологию и почему так важна работа с терминологией?

При использовании терминологического подхода в языковой документации, мы создаем отдельную запись для каждого значения слова. Это может показаться сложным, потому что, если мы создаем терминологическую запись, наша одна статья на рисунке 2 превращается в одиннадцать. Если мы это сделаем, то легко присвоить термину, его синонимам и эквивалентам в нескольких языках или диалектах каждое из значений, то есть точно присвоить эквиваленты каждой идее или понятию, которое он представляет. Такой подход нужен не только для точного перевода, но и для того, чтобы ориентироваться при создании любых новых терминов для понятий, которые нам нужны для процесса оцифровки, или для других отсутствующих понятий, например, в медицине.

Чтобы визуализировать связь между данной идеей или концепцией и термином, который ее представляет, мы часто рисуем треугольник, показывающий (1) *понятие*, которое термин создает в нашем сознании, (2) *объект*, к которому относится термин (иногда называемый *референтом*), или *объект в реальном мире*) и (3) сам термин. Если термин «работает», мы можем использовать его, когда говорим или пишем, и слушатель или читатель может немедленно сформировать в уме ту же идею. Когда мы общаемся более чем на одном языке, важно, чтобы объект (референт) был одним и тем же или очень похожим, и чтобы все на обоих

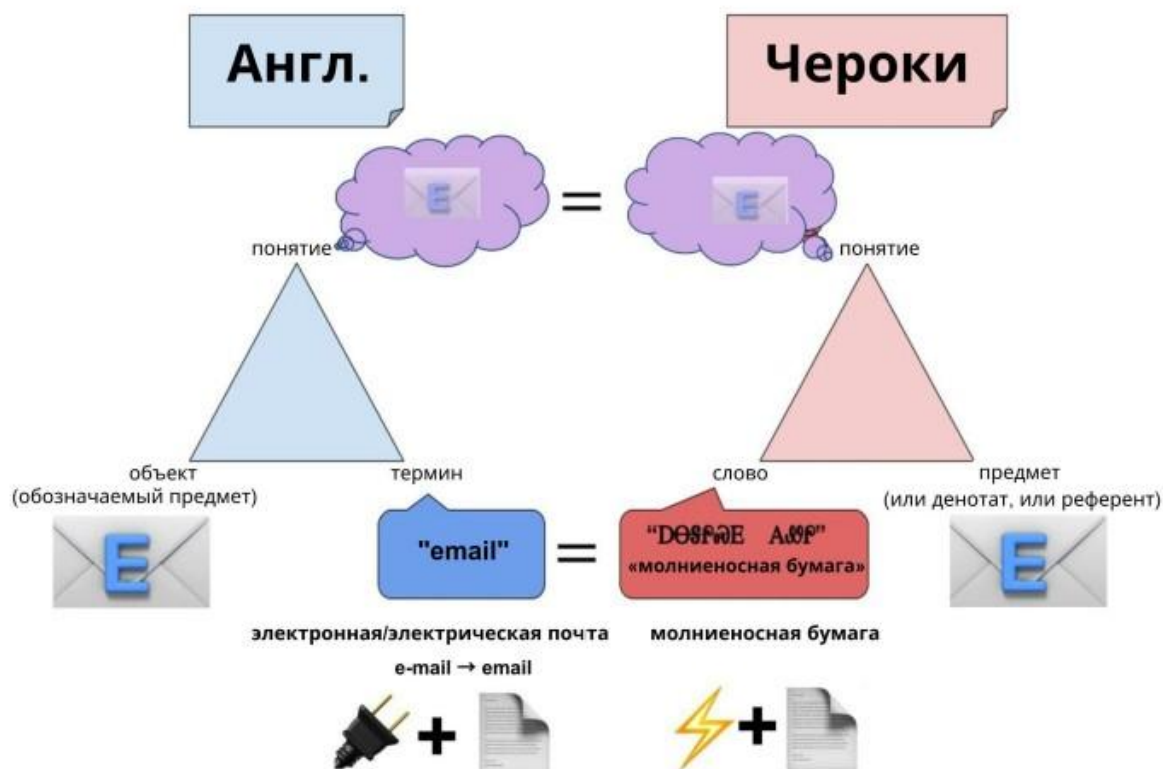
языках понимали одно и то же понятие, которое мы имеем в виду. Только тогда мы можем определить, что термин в языке А адекватно эквивалентен термину в языке Б. Пример семантического треугольника, который используется для представления значения, показан на рисунке 3.

Таким образом, расположение треугольников становится очень полезным, когда мы сравниваем языки. Мы начинаем с признания того, что слово представляет данное понятие. Затем мы ищем эквивалентное понятие в целевом языке. Если для этого понятия нет эквивалентного слова, то это типичное «отсутствующее» слово. Тогда мы можем создать новое или заимствовать слово. В идеале терминологическая работа направлена на выявление существующих в языковом сообществе терминов, а не на заимствование слов из других языков. Термин *tree* является хорошим примером. В языке коренных американцев чероки есть слово *tree*, обозначающее дерево: *ᏒᏐᏃ* (*quigvi* [kʷigēi]). Однако есть много других английских слов, для которых первоначально не было эквивалента в языке чероки. Предположим, мы хотим придумать хороший термин для аналога английского термина *email*. Если надо избежать заимствования терминов, не стоит просто писать слово на языке чероки, которое звучит как «E-MAIL».



**Рисунок 3: «Семантический треугольник»**, где термин «tree» соответствует объектному (референтному) дереву, что вызывает идею или понятие дерева в сознании человека, который слышит или читает термин.

В терминологической работе мы считаем, что независимо от того, имели ли носители английского языка и чероки одинаковый опыт работы с электронными письмами на момент создания терминов, *само электронное письмо является одним и тем же* для них обоих. *Электронная почта* объекта или референта никогда не меняется в этом уравнении. Иногда простые слова, такие как деревья, будут выглядеть для двух говорящих немного по-разному (например, дуб выглядит иначе, чем баобаб), но когда мы называем или определяем наш объект, мы обращаем внимание на общие характеристики двух объектов. В первую очередь это терминологический узел, который, скорее всего, будет отличаться и, вероятно, будет отражать различный опыт говорящих, читающих и пишущих. Сравните английское слово *email*, сокращенное от *electronic mail*, и слово на языке чероки ᎠᎩᎠᎵᎾ ᎠᎩᎵᎾ (*anagalisgv goweli* [anagalisgě goʒeli]), которое дословно переводится как *молниеносная бумага* (lightning paper). Носители английского языка, хоть и видели молнию, ежедневно имеют дело с электричеством, проходящим через электрическую систему (показана здесь электрической вилкой), и они знают, что так доходит до них электронная почта. Конечно, современные носители языка чероки понимают это, но сам язык отстает. Здесь мы имеем конвергенцию концепции электричества, которое проходит по проводу, и электричества в виде молнии, а также скорости и магических свойств электронной почты. И чероки решили использовать естественное слово для обозначения электричества в этом термине, отсюда и *молниеносная бумага*.



**Рисунок 4: Треугольник, показывающий связь между понятиями в двух языках, где совпадающие понятия поддерживают эквивалентные термины для одного и того же объекта-референта.**

Этот подход к сопоставлению термин + идея (понятие) важен, потому что, как уже отмечалось, некоторые термины имеют более одного значения, а некоторые идеи могут быть представлены более чем одним термином, что делает опасным просто пытаться сопоставить термин с термином независимо друг от друга. Этот подход упрощает сопоставление понятий в двух или более языках, позволяет учитывать различные формы в нескольких диалектах и сообщает об альтернативных вариантах написания.

Двухязычные терминологические статьи актуальны потому, что что местные термины часто документируются вместе с их эквивалентами на втором языке, таком как чероки и английский. Обычно мы избегаем заимствования, но если некоторые термины уже были заимствованы и хорошо зарекомендовали себя, вы

можете сохранить их в своем списке терминов или добавить их позже в свою терминологическую базу. Однако, если это возможно, целесообразно создавать собственные термины, используя знакомые слова, а не заимствования. Такой подход помогает носителям языка понимать и управлять оцифровкой своего языка. Преимущества также распространяются на другие области, поскольку создаются новые материалы. Например, в медицине использование знакомых слов помогает медицинским работникам получать доступ к информации, быстро реагировать и общаться, что способствует достижению успешных результатов.

## 6. ДОКУМЕНТИРОВАНИЕ ВАШИХ ТЕРМИНОВ

Списки терминов полезны в начале, но еще вам понадобятся определения, иногда несколько определений, для ваших терминов, и по мере роста вашего списка или списков при необходимости становится все труднее находить информацию в ваших списках. В итоге вы захотите создать *терминологические записи*, особенно если вы хотите задокументировать свои термины вместе с эквивалентами на более чем одном языке. Простая полная терминологическая запись (*модель данных для ввода термина*) примерно может выглядеть так, как показано на рисунке 5.

Типичная информация, включаемая в терминологическую запись, включает следующие элементы данных, обычно называемые *категориями данных*:

- **Предметная область**
- **Определение** понятия и источник, если он у вас есть
- **Термин**
- **Часть речи**
- **Контекст** в примерном предложении, где термин используется типичным образом, плюс источник



Cherokee term	PoS	chr-Definition	chr-Def-Source	chr-Context	chr-Ctx-Source
DƏSŋəE A.ŋŋŋ	noun	DƏSŋəY A.ŋŋŋ: ɔZŋəY DəV.ŋəWŋ DƏSŋəY DhYəJəY EWŋY ɔəY ɬcɔ EJəY DSVi.ŋəJ DəIhAJəY hLGŋəŋ ɔT ɬcɔ Də ɔhAJ JhŋəY ɔtT TJP ɔSLŋŋ	<a href="https://language.cherokee.org">https://language.cherokee.org</a>	ɔ DŋG C DƏSŋəY A.ŋŋŋ ɔSŋəb ɔŋT ɔZŋəY ɔŋŋəLhRY.	Roy Boney <roy-boney@cherokee.org>

English term	PoS	en-Definition	en-Def-Source	en-Context	en-Ctx-Source
email	noun	message distributed by electronic means from one computer user to one or more recipients via a network	<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..69l57j0l512l9.3696j0l15&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..69l57j0l512l9.3696j0l15&amp;sourceid=chrome&amp;ie=UTF-8</a>	I can access my email on my phone or from my computer.	Translation Commons

**Рисунок 6:** Терминологическая запись в виде одной строки в программе для работы с электронными таблицами.

**Предметная область:** Данная категория в вашей записи помогает упорядочить термины в соответствии с темой. Всегда используйте здесь постоянные значения. Например: всегда используйте *Медицинская* (Medical), не меняйте между *Медицинская*(Medical) и *Медицина*(Medicine). Если вы будете последовательны, будет легко отфильтровать ваш ресурс по всем вашим медицинским терминам.

**Определение:** Хорошие определения обычно максимально короткие, но настолько полные, насколько это необходимо. Если термин называет вещь, определение обычно указывает на то, что термин «является» более широким понятием с отличительными характеристиками (например, *волк* (является) *крупным плотоядным животным семейства собачьих*). Для некоторых других терминов вы можете указать «является частью» более широкого понятия (например, *лист* (является) *частью растения, которая прикреплена к растению и поддерживает растение посредством процесса, называемого фотосинтезом*). Глаголы часто определяются пояснительной глагольной фразой. Например, *ходить: двигаться в обычном темпе, поднимая каждую ногу по очереди*. В разделах 7.3 и 7.4 обсуждается создание системных диаграмм понятий. Как только вы соберете

значительное количество терминов, вы можете начать показывать их отношения друг к другу, используя такие графики. В этом процессе потребуется тонкая настройка определений, чтобы они отражали отношения между вашими понятиями.

**Термин:** в начале вашего проекта вы можете просто начать записывать термины, когда вы их слышите. По мере изучения грамматики постарайтесь понять, есть ли для ваших терминов что-то вроде *базовой формы*. Во многих языках это будет наиболее распространенная форма единственного числа существительного, но множественное число может быть другим, а в некоторых языках многие формы по-разному работают в предложении. Если форм много, вам нужно выбрать наилучшие способы их записи на вашем языке. Простую базовую форму часто называют *леммой*. Примеры лемм в английском языке включают «gun» (не «guns» или «running»), «book» (не «books» или «booked»), «red» (не «redden» или «redness») — которые на самом деле являются разными понятиями).

**Часть речи:** в предыдущем абзаце речь шла только о существительных и глаголах, но еще есть небольшие служебные слова, например предлоги, союзы, прилагательные, которые обычно описывают существительные, и наречия, а их описывают глаголы, слова и фразы, кроме существительных. Эти категории называются «частями речи». Эта ситуация меняется от языка к языку, поэтому анализ грамматики вашего языка будет охватывать эти и другие варианты. Часть речи термина и другие грамматические примечания, вероятно, лучше всего записать в ваш одноязычный словарь, а сами термины, независимо от того, какая это часть речи, будут записаны в вашей терминологической базе. Однако уникальные слова, особенно прилагательные и наречия, могут быть записаны в ваш терминологический ресурс.

Следует отметить, что иногда существительные в одном языке могут быть глаголами в другом или наоборот. В английском языке отглагольные существительные часто преобразуются в глаголы действия, а в других языках (например, потаватоми, индейский язык) одушевленные существительные, которые всегда находятся в движении, могут трактоваться как глаголы (например,

слово, обозначающее *реку*). Возможно, вам придется серьезно подумать о том, как *часть речи* работает в вашем языке, и изменить способ записи информации в вашей терминологической базе.

**Контекст:** если вы знаете хороший пример предложения, в котором используется ваш термин, запишите его в это поле. Особенно полезно найти контексты, которые указывают или, по крайней мере, намекают на определение понятия и используют любые идиомы или обороты фразы, обычно используемые с этим термином. Используя термин «книга», например, предложение «Джон покупает книги каждый раз, когда он приезжает в город» не намекает на значение «книги», также как и «Они сели вместе и по очереди читали вслух из сборника рассказов о рыбаках Южно-Яванского моря».

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

## 7. ДРУГАЯ ДЕЯТЕЛЬНОСТЬ

### 7.1 Создание текстового корпуса

Есть ли в вашем языке коллекции существующих текстов? У некоторых языков есть древние письменности, которые только сейчас оцифровываются. Благодаря этому письменность сохраняется и становится более доступной. Другие общины имеют богатые традиции устной «литературы» — рассказов, историй, музыки и поэзии, — которые веками передавались в устной форме, но не были записаны и могут оказаться под угрозой утраты, если число носителей будет сокращаться, а молодые поколения перестанут запоминать тексты. Как отмечалось ранее, вы можете начать с аудиозаписи таких «текстов», а затем расшифровать их для оцифровки. В любом случае, записывая и собирая устную, а так же письменную информацию, вы постепенно соберете текстовый корпус. В целом ваша коллекция может представлять собой *текстовый корпус*. Информация о сборе корпусов, защите

авторских прав и создании архивов содержится в ресурсах [Translation Commons Resources](#).

Существует множество компьютерных инструментов, предназначенных для управления и работы с текстовыми корпусами, включая извлечение терминов. [Инструменты извлечения терминов](#) разработаны специально для идентификации терминов-кандидатов и возможных определений или контекстов, работающих с основными языками. Если вы собрали значительный корпус, вы можете обратиться к программистам, чтобы настроить один или несколько из этих инструментов для работы с вашим языком. При работе с текстами, представляющими культурное наследие вашей группы, важно защищать и уважать убеждения и традиции. Внимательно работайте со своей командой, чтобы обеспечить защиту священных или конфиденциальных материалов, даже если вы работаете над их сохранением для будущих поколений.

## 7.2 Добавление переводов для создания параллельного корпуса

Ценность вашего корпуса сильно возрастает, если вы сможете добавить двуязычный уровень. Это позволит вам расширить свои терминологические ресурсы как за счет человеческого, так и за счет машинного перевода (МТ). Создается этот двуязычный уровень путем перевода текста с вашего языка предложение за предложением на другой язык. Таким образом существует строгое соответствие переведенных предложений и фрагментов предложений, что приводит к созданию памяти переводов, состоящей из так называемых *текстовых сегментов*. Из полученных текстов потом получается «параллельный корпус». Чтобы добавить машинный перевод к вашему набору цифровых ресурсов, сначала нужно создать значительный набор переведенных документов, которые можно использовать для «обучения» вашей автоматизированной системы обнаружению и извлечению последовательно переведенных терминов и сегментов (см. руководство по машинному переводу с нуля).

Первым важным шагом к внедрению решений для перевода является адаптация существующих инструментов автоматизированного перевода (CAT tools) для обработки ваших текстов и правильного отображения вашего языка. Как только переводчики смогут использовать эти инструменты для создания памяти переводов (TM), вы сможете начать сохранять совпадающие сегменты, состоящие из предложений или фраз исходного языка, переведенных на другой язык. Вы можете использовать эти сегменты для сохранения или перестройки целых документов, что позволит вам построить параллельный корпус совпадающих переведенных текстов. По мере роста вашего корпуса вы можете обратиться к крупным поставщикам услуг машинного перевода (MT), чтобы добавить свой язык к языкам, которые поддерживаются их компьютерными системами. Значимые памяти перевода также можно использовать для извлечения не только списков терминов, но и большей части данных, необходимых для полных, содержательных терминологических статей.

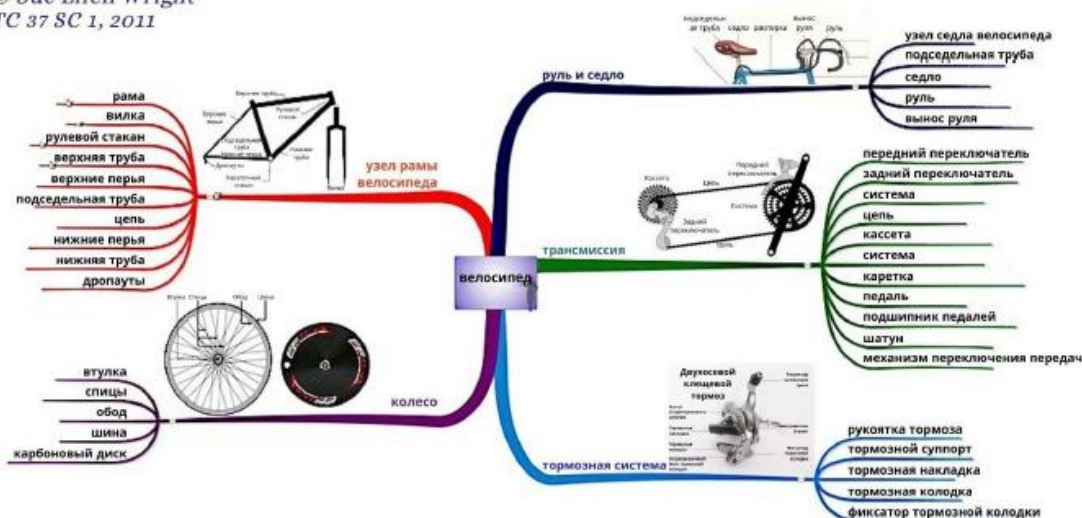
### 7.3 Создание систем общих понятий

Терминологи настоятельно рекомендуют вам систематизировать термины по мере работы над системами понятий. При описании определений мы говорили об отношениях «*является*» и «*является частью*». Спросите себя: «Что такое термин *A*?» Если ответ «*Термин A является разновидностью термина B*», вы можете приступить к построению общей системы понятий. Например, многие люди знакомы с велосипедами. Если бы вам нужно было создать систему понятий, показывающую все виды циклов, которые только можно себе представить, она могла бы выглядеть примерно так, как показано на рисунке 7. Такая система понятий — хороший способ научиться определять понятия. Если вы посмотрите на рисунок, вы можете начать определение, написав: *велосипед — это вид наземного транспортного средства с двумя колесами, которое часто приводится в движение человеком с помощью педалей*.



# 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.**

Рисунок 8: Системы понятий «часть-целое»

Построение системы понятий также является хорошим способом определения отсутствующих терминов, когда вы собираете термины или создаете эквиваленты терминов. Системы понятий также часто показывают, что у термина действительно есть несколько значений, о которых вы изначально не думали.

## 8. СОВМЕСТНОЕ ИСПОЛЬЗОВАНИЕ И ПУБЛИКАЦИЯ ВАШИХ ТЕРМИНОЛОГИЧЕСКИХ РЕСУРСОВ

Как отмечалось выше, чтобы увеличить вашу коллекцию терминологии и сделать ее доступной для более широкой аудитории, лучше всего архивировать ваши данные с помощью системы управления терминологией, которая будет доступна онлайн. Например, программное обеспечение [Terminologue](#), которое было изначально разработано Fiontar & Scoil na Gaeilge для [Foras na Gaeilge](#) для управления Национальной базой данных

терминологии ирландского языка *Téarma* ([téarma.ie](http://tearma.ie)). На сайте Terminologie указано как настроить терминологическую базу и начать вносить данные.

Координируйте свои действия с вашей командой, назначайте редактора и администратора вашей терминологической базы. Сделайте терминологическую базу открытой заинтересованным пользователям для доступа и чтения с компьютера, телефона и других цифровых устройств. Предоставьте гостям и заинтересованным пользователям средства для предложения новых записей и редактирования существующих, но обеспечьте защиту своих данных от нежелательного редактирования или непреднамеренных изменений.

## ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

### Пример проекта:

Ratyegarang – сайт по изучению языкам коренных народов Австралии

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

Данный сайт содержит подробную информацию о том, как придумывать термины. В основе лежит опыт австралийского комьюнити, которое долгое время занимается созданием новых терминов.

### Инструменты:

#### Инструменты для управления текстовыми корпусами

<https://www.Corpus-Analysis.com> предоставляет расширенный список из 261 инструмента для применения в анализе корпуса.

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

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

Существует множество наборов инструментов и постоянно появляются новые. Многие из них доступны бесплатно, а некоторые предоставляют специальные скидки для таких проектов.

#### Управление терминологией:

*Terminologue* — это инструмент 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 извлечения терминов:

<https://termcoord.eu/free-term-extractors/>

## Категории терминологических данных:

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

## Приложения для автоматизированного перевода (CAT Tools), как бесплатные, так и платные:

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

Good Firms «10 лучших 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>

Translate5 Инструмент для перевода с открытым исходным кодом  
<https://www.translate5.net/en/translate5-open-source-translation-system-2>