

**SÉRIE DE ZÉRO À NUMÉRISÉE :**

# **DIRECTIVES DE TERMINOLOGIE**

**UN GUIDE POUR METTRE VOTRE LANGUE EN LIGNE**

**TRANSLATION  
COMMONS**



2022-2032 | INTERNATIONAL DECADE OF  
**Indigenous Languages**

## Directives de terminologie

Auteurs : Sue Ellen Wright, Akil Iyer, Shuto Kato  
Réviseurs et contributeurs : Craig Cornelius, Alaina Brandt,  
Julie Anderson, Tex Texin  
Commercialisation : Leonidas Pappas  
Coordination du projet : Paula Cirilo

Vos retours nous permettent d'améliorer nos guides et  
sont les bienvenus.

Contactez nous à [krista@translationcommons.org](mailto:krista@translationcommons.org)

Cette œuvre est autorisée sous un Attribution Créative  
Commune 4.0. Licence international.  
<https://creativecommons.org/licenses/by/4.0/>

1. INTRODUCTION	4
2. À PROPOS DE CE DOCUMENT	5
3. QU'EST CE QUE LA TERMINOLOGIE ET LE TRAVAIL TERMINOLOGIQUE ?	6
4. LA MÉTHODE IDÉALE À SUIVRE	8
4.1 Êtes-vous prêt à commencer le travail terminologique ?	8
4.2 Avez-vous rassemblé les ressources humaines pour faire le travail ?	9
4.3 Créez des textes et rassemblez les termes existants dans votre langue.	10
4.4 Créer des termes pour les nouvelles idées et concepts qui manquent.	11
4.5 Documentez et maintenez la terminologie dans des glossaires partagés.	11
5. APPLICATION DES DIRECTIVES	12
5.1 Le schéma à suivre	12
5.2 Lexicographie (dictionnaires familiaux) et terminologie (terminographie)	15
5.3 Qu'est-ce-que la terminologie et pourquoi le travail terminologique est-il important ?	16
6. DOCUMENTER VOS TERMES	20
7. AUTRES ACTIVITÉS	24
7.1 Créer un corpus de textes	24
7.2 Ajouter des traductions pour créer un corpus parallèle	25
7.3 Créer des arborescences de concepts génériques	26
7.4 Arborescences de relations partie-tout	27
8. PARTAGER ET PUBLIER VOS RESSOURCES TERMINOLOGIQUES	28
RÉFÉRENCES	30

# 1. INTRODUCTION

[Translation Commons](#) est une organisation à but non lucratif formée d'une communauté de bénévoles qui soutient la numérisation des langues, guide les professionnels de langue, et met à disposition des cours et des ressources pour les industries de la langue.

L'un des principaux programmes à Translation Commons est l'Initiative de numérisation de la langue (I.N.L.), qui cherche à apporter des compétences numériques aux communautés linguistiques qui le désirent. Près de 6 000 langues à travers le monde ont une présence numérique limitée ou inexistante. La I.N.L. fournit des directives qu'une communauté peut suivre pour parvenir à la numérisation de sa langue.

Translation Commons a établi un partenariat avec l'[Année internationale des langues autochtones 2019](#), une initiative de l'UNESCO, pour attirer plus d'attention sur les communautés autochtones et sur la numérisation de leur langues. Une partie de la mission de la I.N.L. est d'aider les locuteurs de langues autochtones et d'autres langues minoritaires à avoir un accès au numérique équitable, pour s'assurer que de telles communautés de langue sont capables de participer aux activités en ligne au niveau mondial et de bénéficier de tous les avantages des applications pour ordinateur modernes, dans leur langue natale. Créer des lignes directrices pour équiper les communautés avec les outils et les connaissances pour numériser leur écriture et mettre leurs langues sur internet leur permettre de faciliter le processus tout en maintenant leur autonomie. En plus de ces directives, Translation Commons fournit des tutoriels, des ateliers et aide les communautés avec la numérisation de leur langue en leurs présentant des experts de l'industrie linguistique, qui les guident à travers le processus de normalisation.

Ce document fait partie d'une série de directives intitulée *De zéro à numérisée*, qui s'adresse de manière globale aux pratiques de numérisation de langue. Les auteurs de ces directives sont des experts dans les technologies de la langue et la linguistique. Le

public visé est n'importe quelle communauté linguistique qui souhaite avoir la capacité d'utiliser leur langue dans un système numérique.

La numérisation ouvre des nouvelles voies de communication aux communautés linguistiques. Voir l'[Annexe des avantages de la numérisation de la langue](#) pour plus de détails sur la manière dont la numérisation d'une langue a des avantages à la fois pour les communautés autochtones et pour le monde en général.

Pour en apprendre plus sur le processus de numérisation de la langue, voir [De zéro à numérisée : Un Guide pour mettre votre langue en ligne](#). Retrouvez plus d'informations sur la I.N.L. et des directives, présentations, vidéos et d'autres documents sur la page [Resources](#) du site internet de Translation Commons.

## 2. À PROPOS DE CE DOCUMENT

Ce document décrit comment documenter la terminologie dans une langue, dans les cas où la documentation existe, est absente ou est insuffisante. Il présente des méthodes pour entrer des mots existants et créer une terminologie nouvelle pour soutenir la communication propre aux domaines spécialisés, pour lesquelles certains mots n'existent pas dans les langues autochtones. La méthodologie recommandée ici conseille sur les divers flux de travail à suivre pour la création ou la mise à jour de la terminologie, y compris l'aménagement, le recueil de données, l'analyse terminologique et la documentation. Cela permet aux communautés autochtones de créer des termes en fonction de leurs besoins. Par exemple, pour des sujets relevant du domaine de la technologie, du domaine médical, politique, légal, ou tout autre domaine qui pourrait être externe aux domaines historiques appartenant à leur culture ou à leur langage. L'élargissement et la mise à jour de ce type de vocabulaire aidera une communauté autochtone à accéder et utiliser des ressources leur facilitant l'intégration dans un environnement numérique.

Certaines langues en danger sont historiquement écrites, tandis que d'autres sont strictement orales. Les langues peuvent aussi être perçues comme des " langues de

famille " qui ne sont pas partagées avec des personnes venant de l'extérieur. Les responsables de la langue devraient documenter les mots existants, en plus de créer de nouveaux mots pour représenter les objets, concepts et idées dont les personnes souhaitant numériser leur langue auront besoin.

Les publics visés par ce guide sont les communautés autochtones qui veulent documenter leur vocabulaire, les experts du domaine, qui soutiennent le travail terminologique dans une ou plusieurs langues et les organisations qui soutiennent les communautés linguistiques dans le processus de numérisation. Le but est d'utiliser la terminologie collectée comme ressource orale, écrite et éventuellement numérique.

Les flux de travail décrites ici peuvent être appliquées au travail terminologique pour n'importe quel domaine ou sujet. Les exemples donnés se basent majoritairement sur la langue anglaise, mais certaines règles générales et modèles peuvent être adaptés à différentes langues et différentes situations. Bien que certaines informations ne s'appliquent pas à toutes les langues. Cependant, ces exemples devraient donner aux lecteurs un bon point de départ, pour identifier les paramètres qui devront être pris en compte dans leur langue lors du travail terminologique.

Ce document reprend les connaissances spécialisées qui sont fondamentales, les meilleures pratiques et les ressources (technologiques ou autres) qui aideront les communautés autochtones à effectuer le travail terminologique dans leurs langues, afin de soutenir les efforts de préservation, de revitalisation et de numérisation des langues.

### 3. QU'EST CE QUE LA TERMINOLOGIE ET LE TRAVAIL

#### TERMINOLOGIQUE ?

Les mots, les abréviations ou même les phrases qui sont utilisés à des fins spécifiques dans les langues sont considérés comme faisant partie de la terminologie de langue de spécialité. Certes, la terminologie de la science et de la technologie est largement

reconnue, mais toutes les activités humaines spécialisées, dans toutes les cultures, utilisent une terminologie spécialisée, que ce soit pour la cuisine, la chasse ou l'expression des valeurs culturelles. Il y a deux buts dans lesquels le travail terminologique est effectué :

- Le recueil de mots et de termes : les mots et termes existants devraient être relevés et documentés. Pour les langues autochtones et sans documentation, cette activité fait partie de la documentation et de l'aménagement linguistique, et il faudrait recueillir tous les mots et termes possibles. Dans ce contexte, il ne faut pas se limiter aux termes des langues de spécialité en tant que tels, sinon attribuer les concepts à leurs domaines de spécialité lorsque c'est pertinent.
- La création de nouveaux termes pour les concepts manquants : dans la mesure du possible, il faudrait utiliser les ressources existantes de la langue ayant été documentées. Pour ainsi de créer de nouveaux termes, lorsqu'il est constaté que certains concepts qu'il faut évoquer n'ont pas encore de termes dans la langue. Les terminologues identifient souvent ces concepts en les comparant avec des concepts similaires et en identifiant un concept qu'ils connaissent dans une langue qui n'a pas été créé dans l'autre. Ces trous sont désignés comme étant des " termes manquants ". Les langues remplissent parfois ces trous avec des emprunts. En anglais, des mots tels que *tsunami* et *Schadenfreude* sont des emprunts. Dans le développement de langues, il est généralement préférable de créer de nouveaux termes en utilisant les ressources existantes plutôt que de les emprunter.

Le travail terminologique est important pour la préservation de la culture et la promotion de l'égalité. Lorsque les langues sont menacées, le travail terminologique peut servir pour créer une archive de la langue et relancer son utilisation. Le gallois et l'hébreu moderne sont des exemples de langues qui étaient menacées de disparition, mais qui sont devenues des langues nationales importantes, notamment grâce à un travail terminologique systématique.

L'objectif du travail terminologique est de recueillir les mots que les personnes utilisent vraiment et de noter le sens communément donné au mot. La collection de mots relatifs aux divers domaines d'activité spécialisés d'une communauté permet aux gardiens de la langue autochtone de participer activement au processus de documentation de leur langue, en préservant l'exactitude des termes et leurs contextes d'utilisation. Comme il a été vu plus haut, le travail terminologique comprend également la création de nouveaux termes lorsqu'une nouvelle technologie est adoptée. Ce qui sera le cas pour de nombreuses langues autochtones en cours de numérisation ou se préparant pour une utilisation sur un appareil numérique. Les termes relevant du numérique, tels que *navigateur* et *outil*, peuvent être empruntés des mots de tous les jours, mais des termes tels que *menu*, *police de caractères* et même *courriel* nécessiteront une réflexion, car leur sens premier peut être inconnu.

## 4. LA MÉTHODE IDÉALE À SUIVRE

Le schéma suivant présente un flux de travail idéal pour le travail terminologiques lié à la numérisation des langues. Toutes les étapes sont importantes pour la réalisation du projet, mais il n'est pas toujours possible de les suivre dans l'ordre précis dont elles sont énumérées. Cette liste est un guide approximatif de votre travail.

### 4.1 Êtes-vous prêt à commencer le travail terminologique ?

- Votre langue possède-t-elle les caractéristiques nécessaires à la numérisation ?  
Par exemple :
- Votre langue a-t-elle un système d'écriture ?
- Votre système d'écriture est-il disponible dans la [norme Unicode](#) pour enregistrer votre terminologie à l'aide d'applications informatiques ?
- Votre système d'écriture a-t-il été configuré pour la saisie de données au clavier et pour l'affichage sur ordinateur et sur appareil numérique ?
- Des règles de grammaire de base ont-elles été déterminées pour votre langue ?

Si la réponse à l'une de ces questions est " NON ", consultez l'onglet [Resources](#) de Translation Commons pour obtenir plus d'informations sur la manière d'aborder ces problèmes. Mais ne vous en arrêtez pas là - même si quelques unes de ces démarches sont en cours, il pourrait y avoir plusieurs autres choses de se faire entre-temps. Certaines des suggestions contenues dans ces lignes directrices peuvent être répétitives, mais elles sont axées sur le processus d'identification et d'enregistrement des termes, ainsi que sur la création de nouveaux termes. Notez bien que vous pouvez commencer à faire les enregistrements audio longtemps avant la finalisation d'une écriture ou d'un système d'écriture et que vous pouvez commencer à utiliser une écriture en évolution, pour recueillir les informations manuellement avant que le processus de numérisation soit terminé.

#### 4.2 Avez-vous rassemblé les ressources humaines pour faire le travail ?

- Avez-vous rassemblé les représentants de votre communauté pour travailler sur votre projet ?
- Avez-vous obtenu le soutien des aînés et des autres responsables du groupe ?
- Quels locuteurs de votre langue ont des connaissances spécialisées pertinentes ? Qui est un bon cuisinier, chasseur, ou a une autre expertise ? Contactez ce cuisinier ou chasseur expert et encouragez-le à vous parler de leur travail.
- Avez-vous trouvé des personnes susceptibles d'apporter un soutien technique pour gérer les tâches informatiques liées au projet ?
- Est-ce-que votre équipe et vous savez comment reconnaître et rassembler des mots et des termes ?
- La constitution de votre équipe peut prendre un certain temps, surtout si vous réunissez des experts dans des [domaines](#) différents, et également le temps d'aider chacun à comprendre vos objectifs. Il pourrait s'agir d'un processus continu en parallèle du projet. Utilisez des tutoriels et autres ressources pour former les personnes à identifier les termes et enregistrer leurs définitions. Si

vous développez votre grammaire et normes d'orthographe, par exemple, et que les personnes doivent apprendre à appliquer ces règles de grammaire de base, ajoutez ces informations à votre programme de formation.

N'oubliez pas de vérifier votre travail au fur et à mesure auprès des membres de votre communauté. Sont-ils d'accord avec vos décisions ? Y a-t-il des divergences de points de vue sur la prononciation ou la manière de représenter les mots dans votre écriture ? Vous pourriez constater que des dialectes émergent pendant le travail. Toutes ces problématiques doivent être examinées afin d'arriver à un consensus dans votre groupe.

#### 4.3 Créez des textes et rassemblez les termes existants dans votre langue.

- Travaillez de préférence sur un seul [sujet](#) spécifique à la fois, mais soyez prêt à documenter tout nouveau sujet qui se présente à vous en indiquant le domaine du sujet au fur et à mesure de votre travail.
- Vérifiez s'il existe déjà des listes de mots, des dictionnaires ou des textes dans votre langue qui traitent du sujet donné et que vous pouvez utiliser pour commencer la collection des termes.
- Parlez aux experts que vous avez identifiés et faites des enregistrements audio de leurs histoires et commentaires.
- Transcrivez vos enregistrements, en commençant par écrire à la main si nécessaire.
- Identifiez les mots et termes importants, puis commencez à faire des listes et à écrire des [définitions](#) ou des explications.
- Si vous documentez les termes de votre espace de vie et que vous ne connaissez pas les termes de certains des objets que vous voyez, essayez de trouver ces termes afin de les inclure dans vos listes.
- Vos termes sont-ils précis ? Vérifiez auprès de plusieurs locuteurs : existe-t-il des variantes de vos termes ? Y a-t-il des dialectes ?
- Documentez les termes pertinents au fur et à mesure que vous avancez.

- Enregistrez vos textes et vos listes de mots au même endroit pour créer un [corpus](#) de textes.
- Une fois que vous avez un grand corpus, il existe des [outils](#) informatiques que vous pouvez utiliser pour organiser votre corpus ou pour documenter vos termes.

#### 4.4 Créer des termes pour les nouvelles idées et concepts qui manquent.

- Y a-t-il des termes dont vous avez besoin qui manquent dans votre langue ?
- Organiser les termes liés dans des systèmes conceptuels afin d'étudier les relations entre les concepts, d'identifier les termes manquants et appuyer la rédaction des [définitions](#).
- Y a-t-il besoin de créer des familles entières de termes, par exemple pour les termes en médecine ou en informatique ?
- Vérifiez avec plusieurs ou même beaucoup de locuteurs si les termes créés ont du sens pour eux.

#### 4.5 Documentez et maintenez la terminologie dans des glossaires partagés.

- Avez-vous décidé comment vous allez vous repérer parmi les termes enregistrés et leur significations ?
- Avez-vous décidé avec quels types de données vous documenterez chaque terme ?
- Avez-vous décidé quel système de gestion terminologique et quel [modèle de données](#) utiliser pour documenter les termes ?
- Chaque concept important a-t-il été documenté dans une collection d'entrées terminologiques, organisée dans une base de données (une base terminologique) ?
- La qualité de chaque entrée d'un concept a-t-elle été vérifiée ?
- La base terminologique a-t-elle été partagée avec les utilisateurs qui en ont besoin ?
- Votre base terminologique décrit-elle ce que vous avez trouvé dans vos recherches ? Cela s'appelle la *terminologie descriptive*.

- Votre base terminologique indique-t-elle comment utiliser les termes ? C'est ce qu'on appelle la *terminologie prescriptive*.
- Les communautés de locuteurs concernées se réfèrent-elles aux concepts dans la base terminologique régulièrement ?
- La terminologie a-t-elle besoin d'être mise à jour ponctuellement ?

## 5. APPLICATION DES DIRECTIVES

### 5.1 Le schéma à suivre

Le flux de travail de la figure 1 donne une vue d'ensemble de ces questions. Gardez en tête que c'est une simplification du processus : vous pourriez vous retrouver à réaliser différentes parties du projet dans un ordre différent ou en parallèle. Chaque projet est différent. Il vaut mieux rester flexible et créatif au fur et à mesure que le travail avance.

Veillez noter que le deuxième élément de la colonne de droite préfère les écritures alphabétiques, où les symboles de l'écriture représentent les sons de la langue sans lien avec leur signification. Une autre option est de travailler avec une écriture logographique, où les caractères de la langue représentent un [concept](#) ou une idée, sans lien à la prononciation du mot ou du terme. L'anglais, le français, et l'arabe sont les exemples d'écritures alphabétiques ; le chinois et les kanji japonais sont des exemples d'écritures logographiques.

Trouver des personnes pour participer au projet devrait être une priorité, et une fois que vous les avez trouvées, il est souhaitable de commencer à travailler avec elles aussitôt que possible. Pendant que vous communiquez avec ces personnes et que vous gagnez leur confiance, vous pourrez alors commencer à faire des enregistrements audio, longtemps avant la finalisation de l'écriture ou du système d'écriture. Vous pouvez aussi commencer à utiliser cette écriture en évolution pour recueillir les informations manuellement, avant que le processus de numérisation soit terminé. Au fil du temps, tandis que de plus en plus de personnes s'impliquent et utilisent des téléphones

portables et autres outils numériques, vous pouvez utiliser les médias numériques pour recueillir des informations auprès d'un groupe grandissant de personnes connaissant la langue.

Ces efforts permettront de gagner un temps précieux et fourniront des informations préliminaires, qui pourront être utilisées pour créer des modèles pour les activités futures, à mesure que le projet progresse. Ce travail préliminaire fournira également des exemples vérifiables de texte, dont vous aurez besoin plus tard lorsque vous demanderez le statut de langue ISO 639, l'enregistrement de écriture ou la certification Unicode. Alors que le projet avance, consultez le flux de travail pour voir si vous devez rattraper certaines étapes que vous n'avez pas pu faire plus tôt.

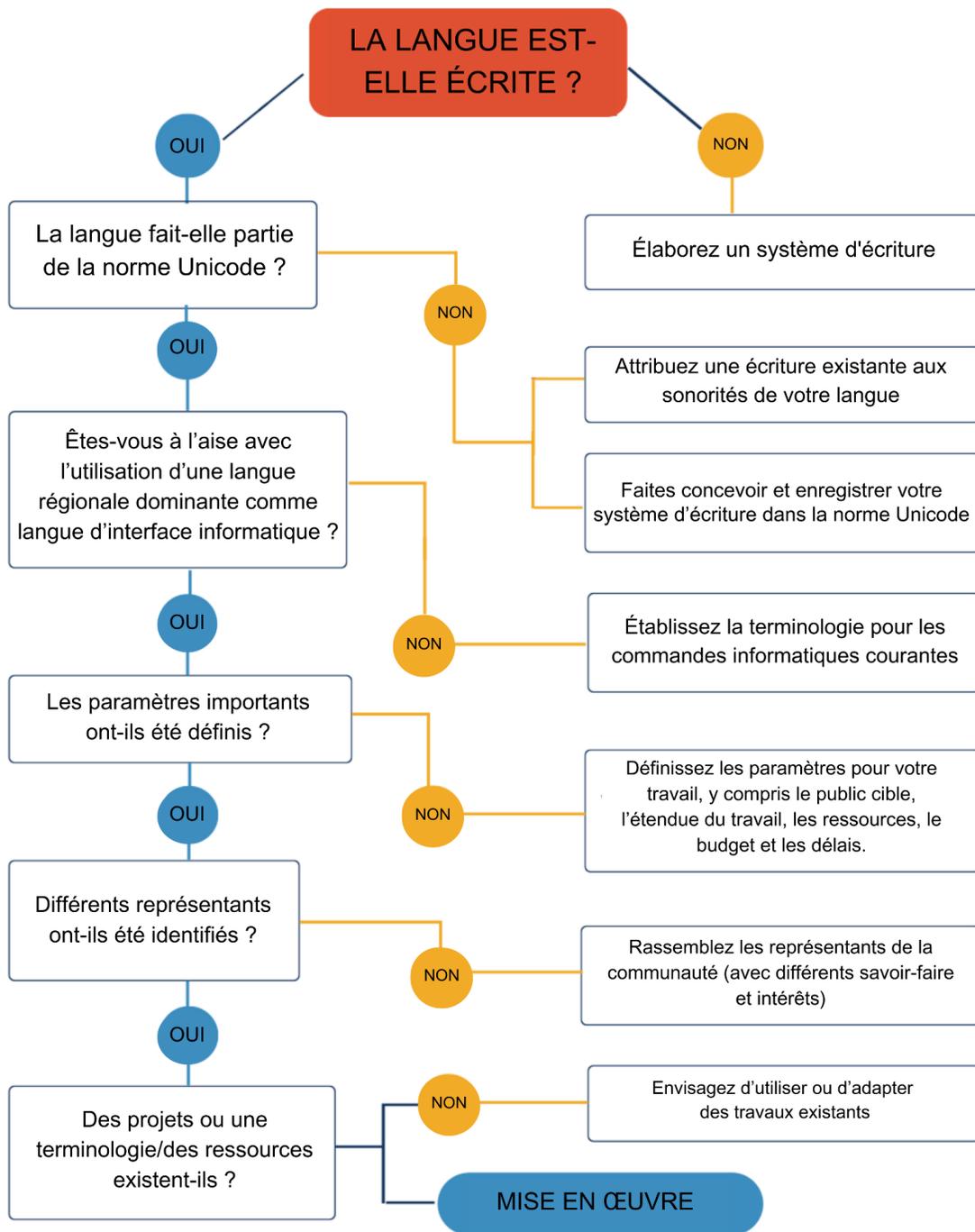


Figure 1 : Flux de travail pour la préparation du projet

## 5.2 Lexicographie (dictionnaires familiaux) et terminologie (terminographie)

Les dictionnaires traditionnels listent des entrées de mots où toutes les significations ([définitions](#)) de ce mot sont incluses dans la même entrée, comme l'illustre la figure 2.

Les dictionnaires lexicographiques sont très utiles quand vous travaillez avec une seule langue, parce que vous pouvez enregistrer toutes les informations pour un seul mot au même endroit. Ces entrées peuvent aussi contenir des informations sur l'utilisation grammaticale et documenter de nombreux petits mots (parfois appelés mots vides dans les collections de corpus). Par exemple les articles, les prépositions, les conjonctions de coordination et d'autres particules linguistiques qui sont utilisés dans la langue courante et la langue spécialisée.

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

1. to make a series of sharp, short sounds in quick succession (produire en rapide succession une série de sons courts et secs)

2. to go or move with such sounds [a wagon rattling over the stones] (aller ou se déplacer en produisant de tels sons)

3. to talk rapidly and incessantly; chatter [often with on: rattle on] (parler rapidement et incessamment ; jacasser)

-vt.

1. to cause to rattle [to rattle the handle of a door] (provoquer un bruit de cliquetis)

2. to utter or perform rapidly (parler ou exécuter rapidement)

3. to confuse or upset; disconcert [to rattle a speaker with catcalls] (rendre confus, blesser ; ébranler)

-n.

1. quick succession of sharp, short sounds (succession rapide de bruits secs et courts)

2. a rattling noise made by air passing through the mucous of a partially closed throat: (crépitements de l'air qui traverse les muqueuses d'une gorge partiellement fermée)

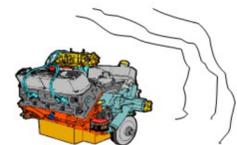
cf. DEATH RATTLE (rôle d'agonie)

3. a noisy uproar; load chatter; engine rattle (brouhaha bruyant ; bruit d'un chargement ; cliquetis de moteur)

4. a series of horny rings at the end of a rattlesnake's tail, used to produce a rattling sound (une série d'anneaux de corne à l'extrémité de la queue d'un serpent à sonnette qui sert à produire le son caractéristique, grelot)

5. a device, as a baby's toy or a percussion instrument, made to rattle when shaken (objet tel qu'un jouet pour enfant [hochet] ou un instrument de percussions [maracas] fabriqué de sorte à faire du bruit lorsqu'il est secoué)

Collocation : to rattle around in a house that is too big for one's needs (être perdu dans une maison trop grandes pour ses besoins)



**Figure 2 : Entrée de dictionnaire (lexicographique) pour le mot anglais *rattle***

Cependant, les entrées lexicographiques deviennent problématiques lorsqu'on essaie de les utiliser comme base pour des dictionnaires bilingues ou multilingues, ou si l'on veut documenter de nombreux mots ou synonymes dans

un dialecte. Si l'on cherche quel sera l'équivalent de chacun des significations de cette entrée dans une autre langue, il est possible que se soit un mot différent. En plus, les différentes orthographes et les synonymes du mot se retrouveront dans plusieurs entrées. C'est pour cela qu'aujourd'hui, la plupart des traducteurs et des autres types de rédacteurs et de chercheurs bilingues, ainsi que l'aménagement linguistique, utilisent ce que l'on appelle " l'approche terminologique " pour entrer les informations lexicales.

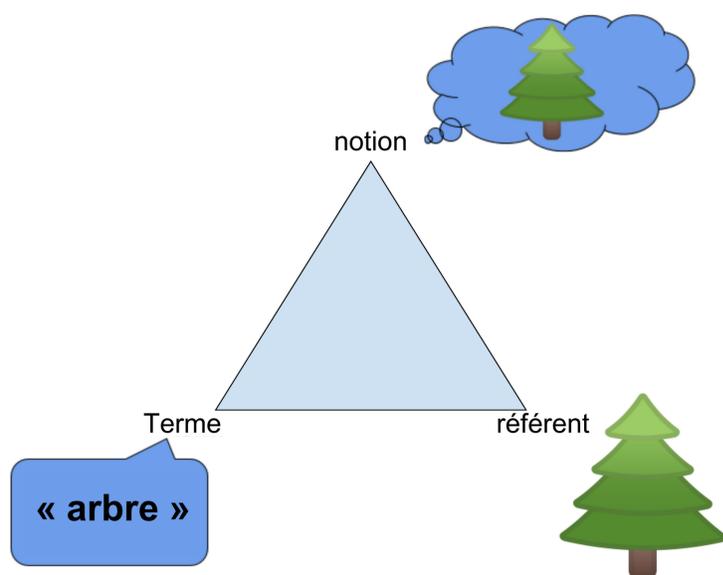
### 5.3 Qu'est-ce-que la terminologie et pourquoi le travail terminologique est-il important ?

Dans l'approche terminologique de la documentation linguistique, une entrée distincte est créée pour chacun des différents sens d'un mot. Cela peut paraître plus compliqué, car l'entrée unique dans la figure 2 deviendrait onze entrées si nous créons des entrées terminologiques. Mais de cette manière, l'attribution des termes des synonymes et des équivalents d'un terme dans plusieurs langues ou dialectes à chacun de leurs sens est facilité. Permettant d'attribuer avec précision des équivalents pour chaque idée ou concept que le terme représente. Cette approche est en fait essentielle, non seulement pour une traduction précise, mais aussi pour se repérer dans le processus de création de nouveaux termes pour les concepts nécessaires à la numérisation ou pour d'autres concepts manquants, par exemple en médecine.

Un triangle est souvent dessiné pour visualiser la relation entre une idée ou un concept donné et le terme qui le représente, illustrant (1) le *concept* que le terme crée dans notre esprit, (2) *l'objet* auquel le terme fait référence (parfois appelé le *référént*, ou *l'objet dans le monde réel*), et (3) le terme lui-même. Si un terme « fonctionne », il peut être utilisé quand on parle ou on écrit, et celui qui nous écoute ou le lecteur peut immédiatement former la même idée dans sa tête. Lorsqu'on communique dans plus d'une langue, il est important que l'objet (le référént) soit le même ou très similaire, et que tout le monde parlant une de ces deux langues comprenne le même concept

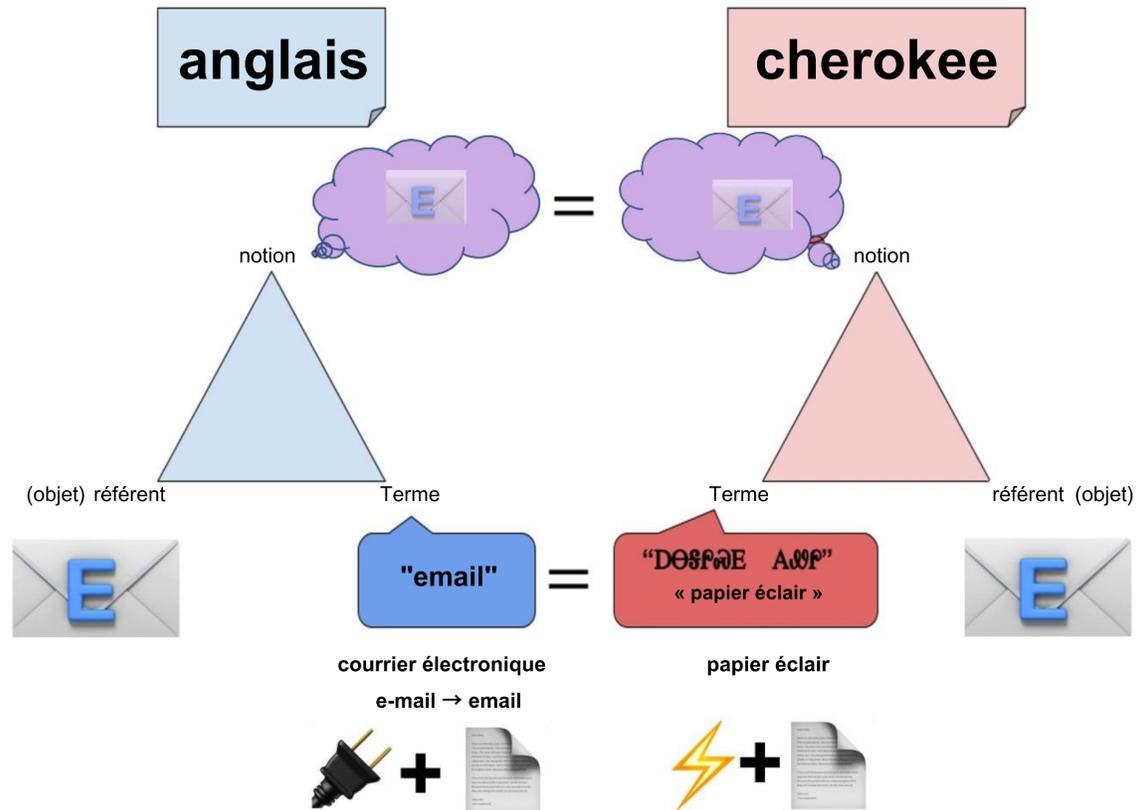
auquel nous pensons. C'est uniquement à ce moment là qu'on peut déterminer si un terme dans une langue A est un équivalent acceptable du terme dans la langue B. Un exemple de triangle sémantique, utilisé pour représenter le sens, est représenté dans la figure 3.

La représentation sous forme de triangle devient alors très utile pour comparer les langues. On commence par identifier le concept donné qu'un mot représente. Ensuite on cherche un concept équivalent dans la langue cible. S'il n'y a aucun équivalent pour ce concept, il s'agit d'une situation typique où un mot est « manquant » et on peut donc créer un nouveau mot ou en emprunter un. Dans l'idéal, le travail terminologique cherche à identifier les termes existants dans une communauté linguistique, plutôt que d'emprunter des mots à des langues de l'extérieur. Le terme anglais *tree* (qui se traduit par arbre) est un bon exemple parce que le cherokee, une langue amérindienne, a un mot équivalent pour *tree* : ᎠᎵᎠᎵ (*quigvi* [kʷigəi]). Mais il y a beaucoup d'autres mots anglais pour lesquels, à l'origine, il n'y avait aucun équivalent en cherokee. Supposons que l'on voulait inventer un bon terme pour le concept représenté par le terme anglais *email*. Pour éviter d'emprunter des termes, on n'épellerait pas tout simplement un mot en cherokee pour qu'il ressemble à « E-MAIL ».



**Figure 3 : "Triangle sémiotique"**, où le terme " tree " correspond à l'objet (le réfèrent), soit un arbre, ce qui déclenche l'idée ou le concept de l'arbre dans l'esprit de la personne qui entend ou lit le terme.

Pour le travail terminologique, on considère qu'indépendamment du fait que le locuteur anglais et le locuteur cherokee aient eu la même expérience des e-mails au moment où les termes ont été créés, *l'e-mail en lui-même* est la même *chose* pour les deux locuteurs. Dans cette relation sémantique, l'objet ou le référent *e-mail* ne changera jamais. Parfois, des objets connus, comme les arbres, auront une apparence un peu différente pour les deux locuteurs (par exemple, un chêne a une apparence différente de celle d'un baobab), mais les caractéristiques communes de ces deux objets sont examinées lorsque ils sont nommés ou définis. C'est principalement le nœud de termes qui est susceptible d'être différent et de refléter les différentes expériences des locuteurs, des lecteurs et des auteurs. Comparez le mot anglais *email* (courriel), venant de *electronic mail* (courrier électronique), et le mot cherokee ᎠᎩᎠᎵᎾᎿ ᎠᎩᎵᎾᎿ (*anagalisgv goweli* [anagalisgã gouweli]) dont la traduction littérale est *papier éclair*. Bien que les locuteurs anglais aient déjà vu des éclairs, ils se servent quotidiennement de l'électricité circulant dans un système électrique (représenté ici par une prise électrique), et ils savent que c'est ainsi que l'e-mail leur parvient. Bien évidemment, les locuteurs modernes du cherokee partagent cette expérience, mais la langue en elle-même n'a pas vécu la même évolution. Ici on observe la convergence du concept de l'électricité qui se déplace à travers un fil électrique et du concept de l'électricité sous la forme d'un éclair, ainsi que la vitesse et les propriétés magiques de l'email. Et le cherokee a choisi d'employer un mot naturel pour l'électricité à travers ce terme, donnant ainsi *papier éclair*.



**Figure 4 : Triangle montrant la relation entre les concepts dans deux langues, où des concepts égaux viennent appuyer les termes équivalents pour le même objet référent.**

Cette approche de faire valoir un terme + une idée (un concept) est importante car, comme nous l'avons vu, certains termes signifient plus d'une chose, et certaines idées peuvent être représentées par plus d'un terme, c'est donc risqué d'essayer de faire correspondre un terme à un autre indépendamment du concept. Cette approche facilite la mise en équivalence d'un concept dans deux langues ou plus, la prise en compte de formes différentes dans les multiples dialectes et l'indication de la présence d'orthographe alternatives.

Une des raisons principales pour laquelle on parle d'entrées terminologiques bilingues est le fait que les termes autochtones sont souvent documentés avec leurs équivalents dans une seconde langue, par exemple en Cherokee et en anglais. Bien qu'il soit généralement mieux d'éviter les emprunts, si certains termes ont déjà été empruntés et

sont bien établis, vous pouvez les conserver dans votre liste terminologique ou les ajouter ultérieurement à votre base terminologique. Si possible, il est tout de même judicieux de créer vos propres termes en utilisant des mots qui vous sont familiers plutôt que de les emprunter. Cette approche permet aux locuteurs de comprendre et de prendre le contrôle sur la numérisation de leur langue. D'autres domaines bénéficient aussi de cette méthode lorsque de nouvelles entrées sont créés. Par exemple, dans le domaine médical, l'utilisation de mots courants aide les professionnels de la santé à accéder aux informations, à réagir et à communiquer rapidement, favorisant l'obtention de résultats positifs.

## 6. DOCUMENTER VOS TERMES

Les listes de termes sont utiles au départ, mais vous aurez également besoin de définitions, parfois plusieurs par terme et à mesure que la ou les listes s'allongent, il devient de plus en plus difficile de trouver des informations dans vos listes lorsque vous en avez besoin. Vous voudrez finalement créer des *entrées terminologiques*, surtout si vous souhaitez documenter vos termes avec leurs équivalents dans plusieurs langues. Une entrée terminologique complète simple (*modèle de données d'entrée de terme*) pourrait ressembler à la figure 5.

Les informations typiques à inclure dans une entrée de terme comprennent les données suivantes, généralement appelées *catégories de données* :

- **Le domaine**
- **La définition** du concept, avec une source si vous en avez une
- **Le terme**
- **La partie du discours**
- **Un contexte**, soit un exemple de phrase où le terme est utilisé de manière caractéristique, en incluant la source

<b>Champ lexical :</b>	communications numériques
<b>Terme cherokee :</b>	DƏSPODE A.ƏP
<b>Nature :</b>	nom
<b>Définition :</b>	DƏSPODY A.ƏP: ƏZPODY DəV.ƏəWƏ DƏSPODY DəYəD.İəDY EWƏY ƏəDY Ɔə E.İəDY DSV.İəD.İ DəD.İH.İəDY İLGLƏSƏ ƏT Ɔə DƏ O.Ĥ.İ.İ.İəDY ƏT T.İP O.ŚL.İP <b>Source :</b> <a href="https://language.cherokee.org">https://language.cherokee.org</a>
<b>Contexte :</b>	Ɔ D.İGG DƏSPODY A.ƏP ƏSƏİB O.PT ƏZPODY ƆƏPəD.LB.H.V. <b>Source :</b> Roy Boney <roy-boney@cherokee.org>
<b>Terme anglais :</b>	email
<b>Nature :</b>	nom
<b>Définition :</b>	message distribué par voie électronique d'un utilisateur d'ordinateur à un ou plusieurs destinataires par l'intermédiaire d'un réseau <b>Source :</b> <a href="https://www.google.com/search?q=define+email&amp;rlz=1C1CHBF_enUS866US866&amp;og=define+email&amp;aqs=chrome..69i57j0i512i9.2230j0i15&amp;sourceid=chrome&amp;ie=UTF-8">https://www.google.com/search?q=define+email&amp;rlz=1C1CHBF_enUS866US866&amp;og=define+email&amp;aqs=chrome..69i57j0i512i9.2230j0i15&amp;sourceid=chrome&amp;ie=UTF-8</a>
<b>Contexte :</b>	Je peux accéder à mes courriels sur mon téléphone ou depuis mon ordinateur. <b>Source :</b> Translation Commons

**Figure 5 :** Structure typique d'une entrée terminologique

Vous pouvez commencer par saisir ces informations dans un fichier texte ou un fichier de traitement de texte, mais il est plus judicieux d'utiliser un tableur pour collecter les entrées de termes. Votre tableur pourrait plus ou moins ressembler à la figure 6. Les deux images dans cette figure sont en réalité une seule ligne dans le tableur. Il existe d'autres agencements plus étroits pour les tableurs, mais la disposition de chaque concept pour les deux (ou toutes) vos langues dans une seule rangée, facilite l'exportation des données vers d'autres systèmes si vous décidez par la suite de créer une base de données. Les tableurs sont utiles pour collecter un petit nombre de termes, mais eux aussi deviennent encombrantes lorsque votre collection de termes s'agrandit. Si vous concevez votre tableur en vous basant sur un modèle de base terminologique, ce sera relativement facile d'exporter ou d'importer vos données.



graphique. Ce processus peut vous amener à affiner vos définitions pour refléter les relations entre vos concepts.

**Terme** : Au début du projet, vous pouvez commencer à enregistrer les termes au fur et à mesure que vous les écoutez. Pendant que vous en apprenez davantage sur la grammaire, essayez de déterminer s'il existe une *forme de base* pour vos termes. Pour une grande partie des langues, ce sera la forme la courante d'un nom, au singulier, mais le pluriel peut présenter des différences, et dans quelques langues, beaucoup des différentes formes peuvent fonctionner différemment dans la phrase. S'il y a beaucoup de formes différentes, il faudra décider des meilleures procédures pour entrer toutes ces formes dans votre langue. La forme de base la plus simple est généralement appelée le *lemme*. Des exemples de lemmes pourraient être : la « course » (pas des « courses » ou « en courant »), « livre » (pas « livres » ou « livré »), « rouge » (pas « rougissez » ou « rougeur », qui sont en fait des concepts différents).

**Partie du discours** : Le paragraphe précédent vient d'évoquer les noms et les verbes, mais il y a aussi les petits mots-outils, comme les prépositions et les conjonctions, les adjectifs qui décrivent habituellement les noms, ainsi que les adverbes qui décrivent les verbes, mots et phrases autres que les noms. Ces catégories s'appellent les « parties du discours. » Cette situation change d'une langue à une autre, alors l'analyse grammaticale de votre langue étudiera ces variations en plus des autres qui pourraient exister. Il est probable que la partie du discours et les autres notes grammaticales d'un terme feront mieux d'être enregistrées dans un dictionnaire monolingue, alors que les termes en eux-mêmes, peut importe la partie du discours, seront enregistrées dans votre base terminologique. Toutefois, les mots uniques, en particulier des adjectifs et des adverbes, peuvent être enregistrés dans votre base terminologique.

Notez bien que parfois les noms dans une langue peuvent être des verbes dans des autres langues ou inversement. En anglais un nom verbal est souvent converti en verbe d'action, et dans d'autres langues (le potawatomi, une langue amérindienne, par exemple) les noms animés qui sont toujours en mouvement sont traités comme s'ils étaient des verbes (le mot pour *rivière*, par exemple). Vous devrez peut-être réfléchir

sérieusement à la manière dont les *éléments du discours* fonctionnent dans votre langue et adapter la façon dont ces informations sont enregistrées dans votre base terminologique.

**Contexte** : Si vous avez un bon exemple de phrase où votre terme est utilisé, enregistrez-le dans ce champ. Trouver des contextes qui indiquent la définition du concept, ou du moins y font allusion, et qui utilisent toutes les expressions idiomatiques ou les tournures de phrases qui sont couramment utilisées avec le terme vous sera particulièrement utile. Si on prend le terme « livre », par exemple, la phrase « Jean achète des livres à chaque fois qu'il est en ville. » ne laisse pas entendre la signification du mot « livre » aussi bien que la phrase : « Ils se sont assis ensemble et ont lu à voix haute, chacun leur tour, des histoires dans un livre sur les pêcheurs de la mer de Java. »

**Note** : L'exemple d'entrée ne montre pas de champ pour les *Notes*, mais elle peut être ajoutée si vous avez des informations additionnelles sur n'importe quel élément de l'entrée.

## 7. AUTRES ACTIVITÉS

### 7.1 Créer un corpus de textes

Votre langue a-t-elle déjà des collections de textes existants ? Certaines langues ont des écritures très anciennes qui sont à peine en train d'être numérisées, ce qui les rendra accessibles et les préservera. D'autres communautés ont une riche tradition de " littérature " orale : récits, histoires, musique et poésie. Qui a été transmise oralement au fil des siècles mais qui n'est pas écrite, et qui peut être menacée de disparition si le nombre de locuteurs diminue et si les nouvelles générations ne mémorisent plus ces informations. Comme indiqué plus haut dans ce document, vous pouvez commencer par l'enregistrement audio de ce type de " textes ", puis les transcrire pour les numériser. Dans tous les cas, vous allez progressivement rassembler un corpus de

textes en enregistrant et en recueillant des informations orales et écrites. Vu dans son ensemble, votre collection peut être composée d'un *corpus de textes*. Vous trouverez des informations sur la collecte de corpus, la préservation des droits d'auteur et la création d'archives sur la page [Resources](#) de Translation Commons.

Il existe une grande variété d'outils informatiques conçus pour manipuler et travailler avec des corpus de textes, ainsi que pour l'extraction de termes. [Les outils d'extraction de termes](#) sont spécialement conçus pour identifier les candidats termes et les définitions ou contextes possibles, en travaillant avec les principales langues parlées. Si vous avez accumulé un corpus important, vous pourrez peut-être faire appel à des informaticiens, pour adapter un ou plusieurs de ces outils à votre langue. Lorsque vous travaillez avec des textes qui sont aussi le patrimoine culturel de votre groupe, Il est toujours important de protéger et de respecter les croyances et les traditions. Travaillez soigneusement avec votre équipe en vous assurant de protéger tout matériel sacré ou confidentiel, même si le travail vise à le préserver pour les générations futures.

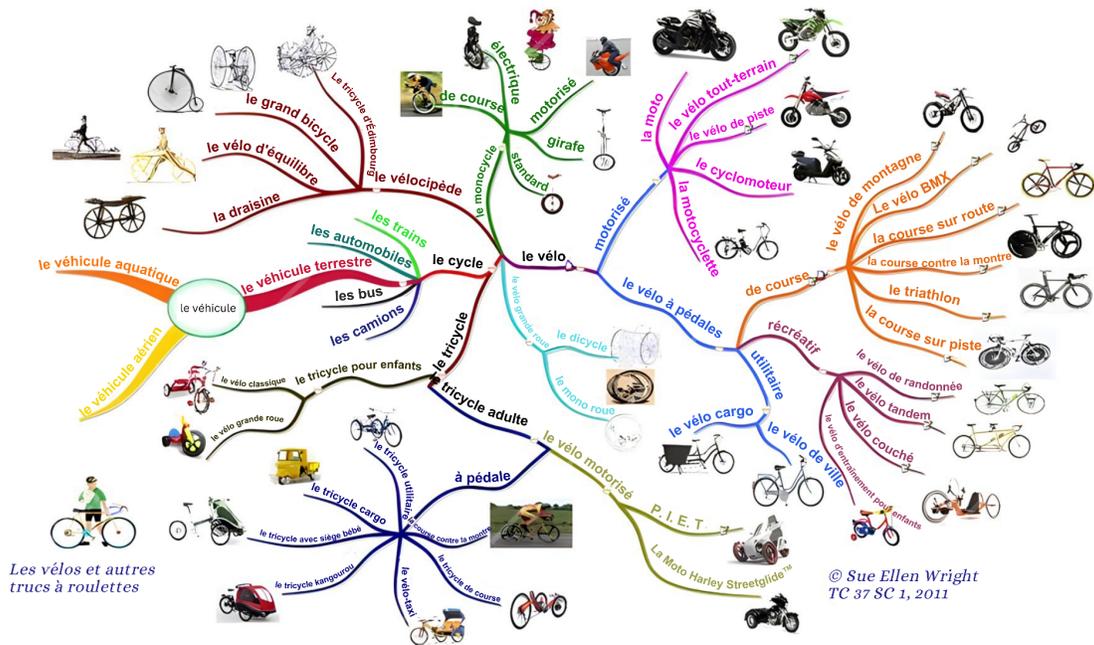
## 7.2 Ajouter des traductions pour créer un corpus parallèle

La valeur de votre corpus est considérablement augmentée si vous pouvez créer une version bilingue, car cela vous permettra d'élargir vos ressources terminologiques grâce à la traduction humaine et la traduction automatique (TA). La création de cette version bilingue, notamment en traduisant le texte dans votre langue, phrase par phrase, dans une autre langue, de sorte qu'il y ait une correspondance stricte entre les phrases et les morceaux de phrases traduits, permet la création d'une mémoire de traduction, constituée de *segments de texte*. Les textes résultants forment alors un « corpus parallèle ». Pour ajouter une TA à vos de ressources numériques, il faut d'abord créer une collection conséquente de documents traduits, qui peuvent être utilisés pour « entraîner » ce système automatisé à découvrir et à extraire des termes et des segments traduits de manière cohérente. (Voir De zéro à numérisée, Guide de traduction automatique.)

La première étape significative dans la mise en œuvre des solutions de traduction consiste à adapter les outils de traduction assistée par ordinateur (TAO) existant déjà, pour qu'ils puissent traiter votre écriture et afficher correctement votre langue. Une fois que les traducteurs humains peuvent utiliser ces outils pour produire une mémoire de traduction (MT), vous pouvez commencer à enregistrer les segments parallèles constitués de phrases ou d'expressions en langue source traduites dans une autre langue. Vous pouvez utiliser ces segments pour enregistrer ou reconstruire des documents entiers, qui permettront de construire un corpus parallèle des textes traduits correspondants. Au fur et à mesure que votre corpus s'enrichit, vous pouvez demander de l'aide aux principaux fournisseurs de TA pour ajouter votre langue aux langues prises en charge par leurs systèmes informatiques. Les MT importantes peuvent également être utilisées pour extraire non seulement des listes de termes, mais aussi une grande partie des données dont vous avez besoin pour créer des entrées terminologiques riches et complètes.

### 7.3 Créer des arborescences de concepts génériques

Les terminologues vous recommandent vivement d'organiser vos termes au fur et à mesure que vous travaillez dans des arborescences. Les relations " *est un*" et " *est une partie de*" ont été évoquées lors de l'explication sur les définitions. Demandez-vous : qu'est-ce-que le terme A ? Si la réponse est que le *Terme A est un type de Terme B*, vous pouvez commencer à construire une arborescence de concepts génériques. De nombreuses personnes savent ce que sont les vélos, par exemple. Si vous deviez créer une arborescence montrant tous les types de deux roues auxquels vous pouvez penser, en anglais, il pourrait ressembler à la figure 7. Cette arborescence de concepts est un bon moyen d'apprendre à définir les concepts. En vous basant sur le dessin, vous pouvez commencer une définition en écrivant : *une bicyclette est une sorte de véhicule terrestre à deux roues, qui est souvent propulsé par l'homme à l'aide de pédales.*



**Un vélo de contre la montre est un type de vélo de course.**  
**Un vélo de course est un type de vélo.**  
**Un vélo est un type de véhicule terrestre.**

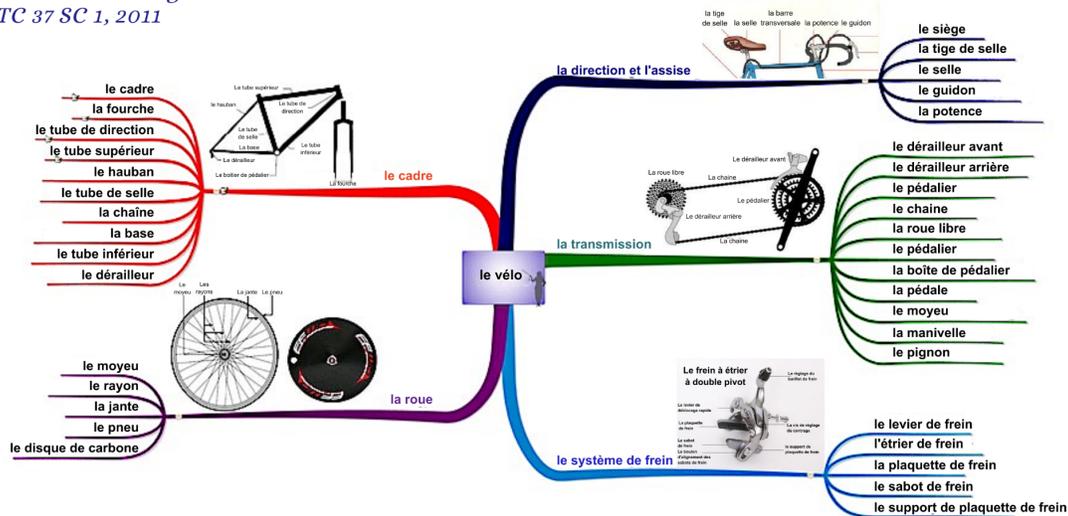
**Figure 7 : Arborescence de concepts génériques**

#### 7.4 Arborescences de relations partie-tout

Vous pouvez également schématiser les concepts à partir de définitions citant les parties qui composent ce concept. Dans l'exemple de la figure 9, la *pédale* fait partie de la *transmission* d'un *vélo*.

# Vélos : pièces et jeux de pièces

© Sue Ellen Wright  
TC 37 SC 1, 2011



*Un jeu de roues fait partie d'un vélo.  
Une roue fait partie d'un jeu de roues.  
Un rayon fait partie d'une roue.*

Figure 8 : Arborescence de relations partie-tout

Dessiner des arborescences aide aussi à identifier les termes manquants lorsque vous collectez des termes ou créez des équivalents de termes. Les arborescences révèlent aussi, qu'en fait, un terme a souvent plusieurs significations auxquelles vous n'aviez pas pensé au départ.

## 8. PARTAGER ET PUBLIER VOS RESSOURCES

### TERMINOLOGIQUES

Comme nous l'avons déjà mentionné, la meilleure façon d'accroître votre collection terminologique et de la rendre accessible à un plus grand nombre de personnes est d'archiver vos données à l'aide d'un système de gestion terminologique, qui sera facilement accessible sur le Web. Par exemple, le logiciel [Terminologue](#), qui à l'origine a été développé par Fiontar & Scoil na Gaeilge pour [Foras na Gaeilge](#) afin de gérer *Téarma*, la base terminologique nationale pour

la langue irlandaise (National Terminology Database for Irish, [téarma.ie](http://tearma.ie)). Le site Web de Terminologie donne des explications sur comment créer des bases terminologiques et comment commencer à saisir les données.

Communiquez avec votre groupe pour décider qui sera le principal responsable de l'édition et de l'administration de votre base terminologique. Faites en sorte que les utilisateurs intéressés puissent accéder à la base de données sur leurs ordinateurs, téléphones et autres appareils numériques. Et donnez les moyens aux invités et aux personnes intéressées de suggérer de nouvelles entrées et de modifier les entrées existantes sur la plate-forme, mais protégez bien vos données contre toute modification non souhaitée ou involontaire.

## RÉFÉRENCES

### **Exemple de projet :**

Patyegarang, un site pédagogique pour les langues autochtones australiennes

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

Ce site web fournit d'excellentes informations sur la manière de créer des termes, basé sur les connaissances d'une communauté australienne qui a beaucoup d'expérience dans la création de nouveaux termes.

### **Outils :**

#### **Outils de gestion des corpus de texte**

<https://www.Corpus-Analysis.com> fournit une liste très complète de 261 outils à utiliser pour l'analyse de corpus.

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

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

Il existe de nombreuses collections d'outils différents et la sélection est en constante évolution. Beaucoup d'entre eux sont disponibles gratuitement et certains peuvent accorder des remises exceptionnelles pour des projets comme ceux-ci.

### **Gestion de la terminologie :**

*Terminologue* est un outil de gestion terminologique open-source. Le logiciel est développé et maintenu par le [Gaois research group](#) à [Fiontar & Scoil na Gaeilge, Dublin City University](#). Le logiciel est protégé par les droits d'auteur de [Dublin City University](#) et est disponible sous [licence MIT](#) en open-source. Le développeur principal est [Michal Boleslav Měchura](#). Téléchargez le logiciel à partir de [ce](#) site si vous souhaitez installer votre propre version de *Terminologue*.

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

### **Outils d'extraction de termes :**

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

### **Catégories de données terminologiques :**

Il existe beaucoup d'autres catégories de données qui peuvent être utilisées dans les entrées terminologiques. Vous en trouverez toute une liste, avec des définitions et des exemples de leur utilisation sur <https://www.datcatinfo.net>.

### **Outils d'aide informatique à la traduction (Outils TAO), gratuits et payants :**

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

Good Firms. " Le Top 10 des logiciels de traduction assistée par ordinateur Open Source et gratuit "

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

Translate5 outil de traduction Open Source.

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

**Linguistes:**

Yuna-Aisha Guillamot  
Annette Mitchell  
Camille Krieg

**Chefs de projets de localisation:**

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

**Ingénieures en localisation:**

Luiyi Yang  
Jiaqian Wang  
Yiwen Wang