

STATION RELAIS NATIONALE ELRI

GUIDE DE L'UTILISATEUR

Ce document fournit des lignes directrices à l'intention des utilisateurs des Stations Relais Nationales développées dans le cadre du projet ELRI (European Language Resource Infrastructure).¹

Niveau de diffusion	Public
Version	V2.1
Date	2019/03/07
Auteur(s)	Thierry Etchegoyhen
	Consortium ELRI



The sole responsibility of this publication lies with the author. The European Union is not responsible for any use that may be made of the information contained therein.

¹www.elri-project.eu



Table des matières

1	Intro	oduction	3
2 Rejoindre une Station Relais Nationale			
	2.1	Inscription des utilisateurs	4
	2.2	Autorisation de l'utilisateur	5
3	Accè	es aux services d'ELRI	5
4	Utili	sation des services d'ELRI	6
	4.1	Qu'est-ce qu'une ressource linguistique ?	е
	4.2	Que peut-on télécharger ?	7
	4.3	Que peut-on déposer ?	8
	4.4	Qui partage avec qui ?	8
5	Prod	édure pas à pas pour les utilisateurs	9
	5.1	Partager des ressources linguistiques	9
	5.2	Téléchargement de ressources linguistiques	. 13
6	Con	tacter l'équipe de la SRN	. 16



1 Introduction

Les Stations Relais Nationales (SRN) du projet ELRI sont des applications web conçues pour offrir des services qui facilitent la préparation et le partage des ressources linguistiques (RL) à différents niveaux : avec toutes les organisations publiques enregistrées au niveau national, entre des groupes spécifiques d'organisations enregistrées, avec la Commission européenne ou en tant que données ouvertes.

Le réseau est limité aux organismes publics d'un État membre donné. En d'autres termes, seul le personnel d'une entité publique peut s'inscrire en tant qu'utilisateur de la SRN dans son État membre. Dans une SRN, les ressources ne sont donc partagées qu'entre les organismes publics enregistrés. Les ressources que les propriétaires sont disposés à partager avec un public plus large peuvent choisir de les partager avec des groupes plus importants et ces ressources seront placées dans des répositoires dédiés tels que ELRC-SHARE ou les portails européens de données ouvertes.

Chaque SRN est déployée dans un État membre de l'Union européenne² et elle est maintenue par une équipe dédiée dans l'État membre. Cette équipe supervise la gestion des utilisateurs et la bonne gestion des ressources, tout en servant également d'interlocuteur pour les utilisateurs des services ELRI dans ce pays et pour les langues concernées.

Dans ce document, les utilisateurs peuvent trouver des informations sur l'adhésion et l'utilisation de la Station Relais Nationale dans leur État membre.

La page d'accueil de la SRN en France est illustrée à la Figure 1.

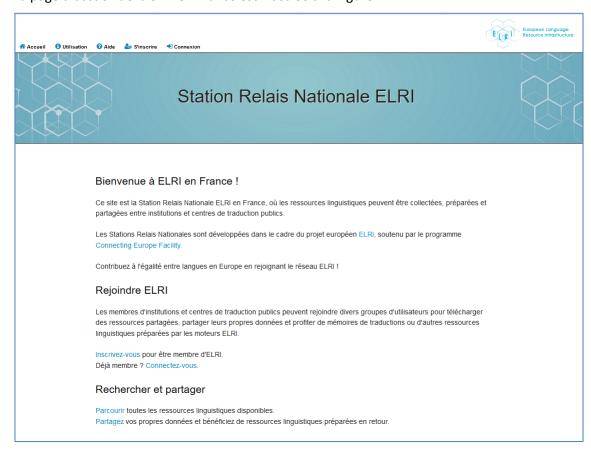


Figure 1 Page d'accueil de la Station Relais Nationale en France

²Actuellement : Espagne, France, Irlande, et Portugal.



2 Rejoindre une Station Relais Nationale

Une Station Relais Nationale ELRI n'est accessible qu'aux utilisateurs enregistrés et autorisés des organismes publics de l'État membre dans lequel la SRN est déployée. L'adhésion à une SRN est donc un processus en deux étapes :

- 1. Les utilisateurs souhaitant rejoindre le réseau peuvent demander la création d'un compte via la page d'inscription de la SRN.
- 2. L'accès est accordé par les membres de l'équipe dédiée de la SRN aux utilisateurs qui travaillent dans une organisation publique.

Les sections 2.1 et 2.2 ci-dessous fournissent de plus amples détails sur chacune de ces deux étapes.

2.1 Inscription des utilisateurs

Les nouveaux utilisateurs peuvent s'inscrire en ligne via la page **S'inscrire** indiquée à la Figure 2.

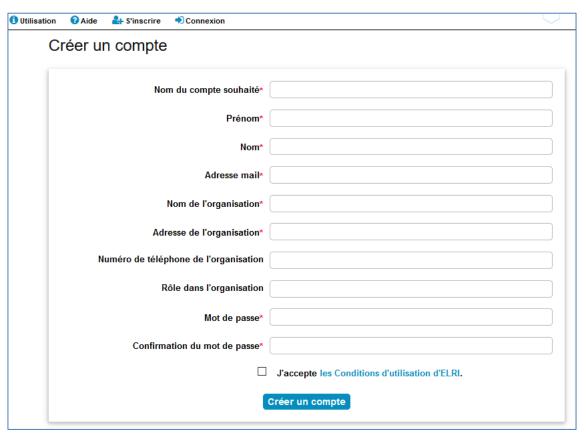


Figure 2 Page d'inscription

Après avoir fourni les informations obligatoires demandées³, les utilisateurs doivent accepter les Conditions d'utilisation d'ELRI en cochant la case correspondante avant de cliquer sur le bouton *Créer un compte*.

Une fois la demande d'inscription envoyée, elle sera évaluée et finalement validée comme décrit à la section 2.2.

³Signalées par un astérisque rouge.



2.2 Autorisation de l'utilisateur

Les demandes d'inscription parviennent à l'équipe ELRI dédiée responsable en France. Le personnel d'ELRI pour cet État membre examinera la demande et, après avoir vérifié que l'utilisateur appartient à une institution publique de l'État membre, validera le compte demandé.

Après validation, l'utilisateur recevra une notification par courriel, comme indiqué dans la Figure 3, et sera en mesure d'accéder aux services d'ELRI proposés par la Station Relais Nationale.

```
Cher/Chère Utilisateur Dupont,

Nous avons activé votre compte utilisateur ELRI "Utilisateur", bienvenue chez ELRI ! Vous pouvez maintenant accéder à nos services en vous connectant à notre site web https://elri.elda.org.

Bien cordialement,

L'équipe ELRI
```

Figure 3 Courriel de notification de validation d'inscription

3 Accès aux services d'ELRI

Seuls les utilisateurs enregistrés et autorisés ont accès aux services fournis par l'ELRI, à savoir la navigation, le téléchargement et le partage de ressources linguistiques. Avant d'accéder à ces services, les utilisateurs doivent donc d'abord se connecter au système via la page de **Connexion** indiquée à la Figure 4 avec le nom d'utilisateur et le mot de passe qu'ils ont choisis lors de leur inscription.



Figure 4 Page d'authentification de l'utilisateur

Une fois connectés, les utilisateurs ont accès aux options du menu indiquées à la Figure 5.



Figure 5 Options du menu pour les utilisateurs connectés

À partir de ce menu, les utilisateurs ont accès aux services suivants :

- Accueil: lien vers la page d'accueil.
- <u>Parcourir</u>: lien vers la page où les utilisateurs peuvent rechercher et télécharger des RL (voir la section 5.2).



- <u>Partager</u>: lien vers la page où les utilisateurs peuvent partager des RL (voir la section 5.1).
- <u>Utilisation</u>: lien vers les instructions d'utilisation du système.
- <u>Conditions d'utilisation</u> : lien vers les Conditions d'utilisation d'ELRI.
- <u>Aide</u> : lien vers une page d'aide avec des coordonnées de contact et de la documentation de support.
- <u>Nom d'utilisateur</u>⁴: lien vers une page contenant les informations de profil de l'utilisateur actuellement connecté.
- Se déconnecter : déconnecte l'utilisateur du système et ferme la session.

4 Utilisation des services d'ELRI

Par l'intermédiaire de ses Stations Relais Nationales, ELRI fournit essentiellement les moyens de préparer et de partager des ressources linguistiques. Dans les sections qui suivent, nous expliquons ce que sont les ressources linguistiques, comment les utilisateurs peuvent contribuer et comment le partage fonctionne dans une SRN.

4.1 Qu'est-ce qu'une ressource linguistique?

Le terme *ressource linguistique* englobe toutes les données pertinentes pour fournir des services linguistiques dans une langue ou un groupe de langues donné. Cela comprend les principaux types de ressources suivantes :

- Corpus: Il s'agit du type principal de RL et il s'agit essentiellement d'un jeu préparé de documents contenant du texte dans une ou plusieurs langues. Un corpus peut être monolingue, c'est-à-dire contenant du texte dans une seule langue dans un seul fichier, ou bilingue, c'est-à-dire contenant des textes traduits dans deux langues et dans des fichiers séparés. Les corpus peuvent être créés à partir de collections de documents dans différents formats (MS Word ou RTF, par exemple), en extrayant et en traitant le contenu des documents.
- Mémoire de traduction : l'une des ressources les plus utiles à la fois pour les traducteurs et les services de traduction automatique est la mémoire de traduction, qui est un document structuré avec l'extension de fichier TMX standard, où les traductions entre les langues sont appariées en termes d'unités de traduction (généralement des phrases alignées par paire, c'est-à-dire une source/un original et son équivalent traduit dans la langue cible). Un exemple de contenu TMX est fourni en Figure 6.
- <u>Fichiers d'échange terminologique</u>: les correspondances terminologiques dans différentes langues peuvent également être stockées de manière structurée, généralement dans des fichiers TBX similaires aux fichiers TMX décrits ci-dessus, mais contenant des termes équivalents appariés au lieu de phrases appariées.

Il existe d'autres types de ressources linguistiques, telles que les descriptions grammaticales d'une langue ou les thésaurus. Toutefois, pour des raisons pratiques, les trois types de RL décrits ci-dessus sont les plus utiles pour améliorer les services de traduction.

⁴ Utilisateur dans l'exemple.



Les mémoires de traduction et les fichiers terminologiques, en particulier, sont des ressources extrêmement importantes pour les services de traduction : les traducteurs humains s'appuient dessus pour améliorer leur productivité, en étant capables d'extraire et de réutiliser du matériel déjà traduit, tandis que les systèmes modernes de traduction automatique apprennent à traduire à partir des traductions humaines stockées dans des corpus bilingues et des mémoires de traduction.

Fournir des ressources linguistiques de qualité est donc un élément essentiel de la construction d'une Europe véritablement multilingue et ELRI vise à fournir des services qui facilitent la création et le partage de ces ressources.

```
[...]
  <tu tuid="1" creationdate="2018-09-07T12:01:58Z" creationid="ELRI">
  <tuv xml:lang="fr">
   <seg>Ceci est un exemple.</seg>
   <tuv xml:lang="en">
    <seg>This is an example.</seg>
   </tuv>
  </tu>
  <tu tuid="2" creationdate="2018-09-07T12:01:58Z" creationid="ELRI">
   <tuv xml:lang="fr">
    <seg>Une mémoire de traduction contient des unités de traduction.</seg>
   <tuv xml:lang="en">
    <seg>A translation memory contains translation units.</seg>
   </tuv>
  </tu>
[...]
```

Figure 6 Exemple de contenu TMX

4.2 Que peut-on télécharger ?

Les ressources partagées sont présentées dans un format préparé, créé automatiquement par les moteurs d'ELRI à partir des données déposées par les utilisateurs. En d'autres termes, les données originales fournies par les utilisateurs (par exemple, un jeu de documents en plusieurs langues) ne sont pas partagées telles quelles : elles subissent une série d'étapes de traitement conduisant à la création de ressources linguistiques structurées qui sont ensuite partagées conformément aux autorisations spécifiées par les utilisateurs qui ont mis en ligne les données d'origine.

Ces RL structurées se présentent sous les formats suivants :

- <u>Mémoires de traduction</u>: fichiers au format TMX standard, où les traductions sont appariées au niveau de la phrase, à la suite d'un processus d'alignement.
- <u>Fichiers monolingues</u>: lorsqu'un document dans une langue donnée n'a pas d'équivalent de traduction dans une autre langue, les moteurs d'ELRI continuent à produire un fichier traité à partir du fichier original, le contenu du fichier étant divisé en phrases distinctes, avec une phrase par ligne.
- <u>Fichiers d'échange terminologique</u> : fichiers au format TBX standard, dans lesquels des termes équivalents dans différentes langues sont appariés.



Toutes les ressources téléchargeables sont présentées dans un fichier compressé (archive.zip) qui contient un ou plusieurs fichiers dans le ou les formats spécifiés ci-dessus.

Les étapes relatives à la mise en ligne de ressources sont décrites en détail à la section 5.1.

4.3 Que peut-on déposer ?

Comme indiqué précédemment, les Stations Relais Nationales intègrent les moteurs de traitement des données d'ELRI. La fonction de ces moteurs consiste à créer des ressources linguistiques structurées à partir de données brutes fournies par les utilisateurs dans différentes langues et dans un grand nombre de formats, y compris MS Word doc(x), ODT ou RTF.⁵

Les utilisateurs peuvent ainsi simplement déposer différents documents dans une ou plusieurs langues et laisser les moteurs d'ELRI préparer des ressources structurées en arrière-plan. Une fois que les ressources ont été préparées et validées par l'équipe d'ELRI dédiée en France, elles seront ensuite publiées sur le site web et seront visibles conformément aux restrictions de partage spécifiées par l'utilisateur qui a fourni les ressources.

Les principaux scénarios concernant le dépôt de données sont les suivants :

- Charger une collection de documents avec le même contenu traduit dans deux ou plusieurs langues différentes. Les moteurs d'ELRI produiront ensuite des mémoires de traduction à partir du contenu traduit en appariant automatiquement les documents et les phrases des documents originaux.
- Charger des mémoires de traduction existantes. Dans la pratique, les mémoires de traduction peuvent contenir des erreurs, telles que des unités de traduction dans la mauvaise langue ou des phrases mal alignées. Dans ce cas, les moteurs d'ELRI offrent la possibilité de nettoyer automatiquement les mémoires de traduction existantes.
- Charger des fichiers terminologiques ou d'autres ressources. Les utilisateurs peuvent également fournir des fichiers qui ne nécessitent pas de traitement automatique. Ces fichiers seront également examinés et validés par l'équipe d'ELRI avant de publier et de partager les ressources.

Les étapes nécessaires au dépôt des ressources linguistiques sont décrites en détail à la section 5.1.

4.4 Qui partage avec qui?

ELRI a pour objectif de fournir des moyens souples de partage des données, tout en reconnaissant que les restrictions de partage peuvent devoir varier pour des ressources spécifiques.

Le partage via une SRN se fait par le biais de **groupes**, où les utilisateurs peuvent consulter et télécharger uniquement les ressources partagées avec un groupe auquel ils appartiennent. Il existe trois groupes différents auxquels les utilisateurs d'une SRN appartiennent par défaut :

- OrganisationsNationales: Ce groupe comprend tous les utilisateurs enregistrés de la SRN et les ressources partagées avec ce groupe seront donc accessibles à tous les utilisateurs enregistrés de la SRN.
- OrganisationsNationales+CommissionEuropéenne : Ce groupe comprend tous les utilisateurs enregistrés de la SRN et de la Commission européenne, qui peuvent alors

⁵ La liste complète des formats de fichiers acceptés est disponible sur la page *Partager* de la SRN.



utiliser les ressources partagées pour améliorer les services de traduction automatique fournis librement aux administrations publiques des États membres de l'UE eTranslation). Les ressources partagées avec ce groupe seront donc accessibles à tous les utilisateurs enregistrés de la SRN et seront également partagées avec la Commission européenne en étant placées sur le répositoire ELRC-SHARE.⁶

 Données ouvertes: Ce groupe comprend tous les utilisateurs enregistrés de la SRN et tous les utilisateurs du portail gratuit des données ouvertes de l'Union européenne.⁷
 Les ressources partagées avec ce groupe seront donc placées dans le répositoire de données ouvertes, en plus de la SRN.

Ces groupes par défaut sont toujours disponibles pour les contributeurs de données et visent à couvrir les cas les plus fréquents de partage des ressources. Si des besoins de partage différents surviennent pour des ressources spécifiques, les utilisateurs peuvent demander la création *ad hoc* de groupes spécifiques en contactant l'équipe ELRI en France.

À titre d'exemple simple, les utilisateurs d'une organisation publique donnée (appelée *OrganisationX*) peuvent avoir besoin de restreindre le partage d'une ressource aux membres de l'OrganisationX *et* de la Commission européenne seule, afin de contribuer spécifiquement à l'amélioration des services de traduction automatique de l'Union européenne, sans partage supplémentaire en raison de restrictions de distribution spécifiques.

Cela peut être réalisé en demandant la création d'un groupe spécifique réunissant les deux entités, par exemple *OrganisationX+CommissionEuropéenne*. Les utilisateurs de l'*OrganisationX* pourront alors partager leur ressource et sélectionner ce groupe spécifique pour répondre au niveau de partage qu'ils souhaitent.

5 Procédure pas à pas pour les utilisateurs

Les deux sections ci-dessous fournissent un guide par étapes pour les utilisateurs souhaitant accéder aux ressources linguistiques ou y contribuer. Pour accéder aux services, les utilisateurs doivent d'abord se connecter au système, comme décrit à la section 3.

5.1 Partager des ressources linguistiques

Pour contribuer avec leurs propres ressources, les utilisateurs doivent cliquer sur l'option **Partager** dans la barre de menu, qui ouvre la page où les informations sur la ressource peuvent être fournies. Le haut de la page est affiché dans la Figure 7.

⁶https://elrc-share.eu/

⁷https://data.europa.eu/euodp/en/home



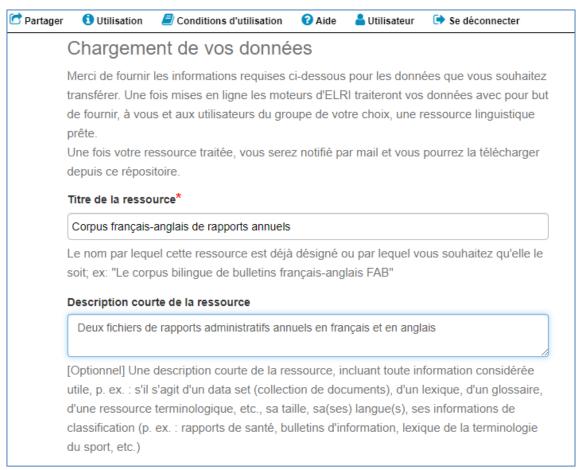


Figure 7 Informations de base sur la ressource

Sous « Titre de la ressource » veuillez entrer un nom défini par l'utilisateur pour la ressource, et sous "Description courte de la ressource » veuillez fournir une description pour la ressource ; par exemple :

« Une version française et anglaise d'un communiqué de presse envoyé le 12/5/18 concernant le lancement d'une nouvelle initiative »

Plus bas dans la page, l'utilisateur doit spécifier la ou les langues de la ressource (cette action est obligatoire, comme l'indique l'astérisque rouge). Dans cet exemple simple, les fichiers contiennent des textes en anglais et en français et ces deux langues doivent être sélectionnées, comme indiqué dans la Figure 8. En outre, l'utilisateur a la possibilité de sélectionner également le ou les domaines concernés dans la liste des options fournies sous la rubrique correspondante (« Autre/Inconnu »).



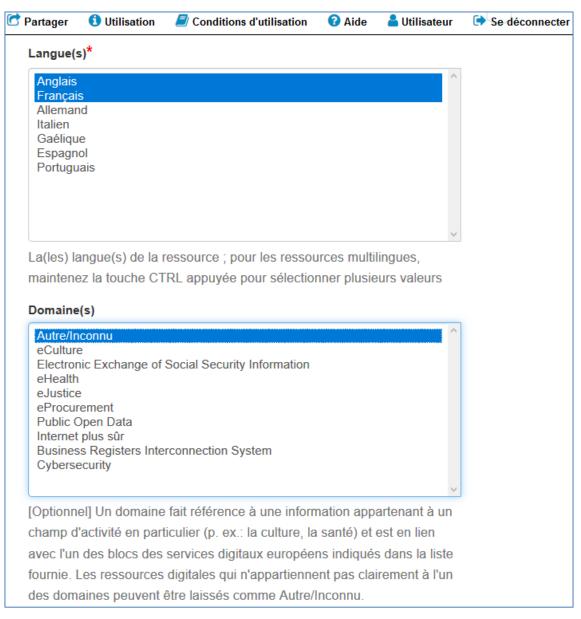


Figure 8 Langues et domaine

Enfin, le bas de la page permet à l'utilisateur de spécifier les conditions de licence associées à la ressource, de charger les fichiers et de sélectionner le ou les groupes avec lesquels partager la ressource, comme le montre la Figure 9. Il s'agit d'actions obligatoires, comme l'indiquent les astérisques rouges.



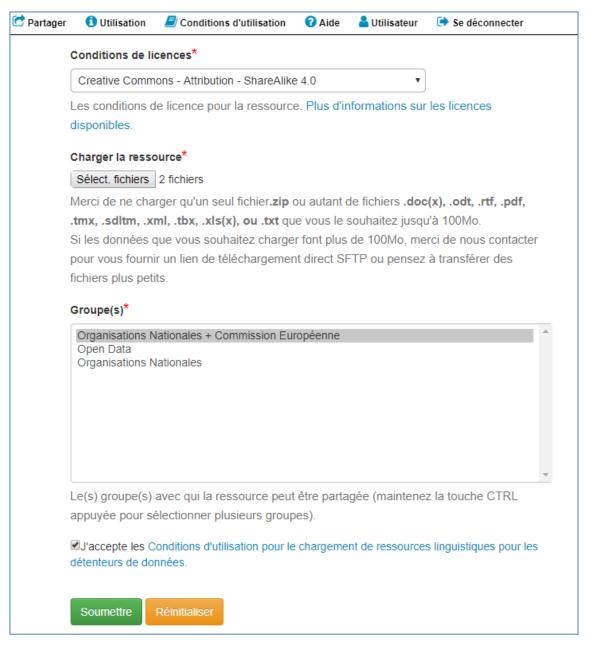


Figure 9 License, chargement de fichier et sélection de groupe

Après avoir accepté les Conditions d'utilisation liées à la mise en ligne de la ressource et cliqué sur le bouton **Soumettre**, les fichiers sont chargés. Si le chargement est réussi, le message indiqué en Figure apparaîtra.

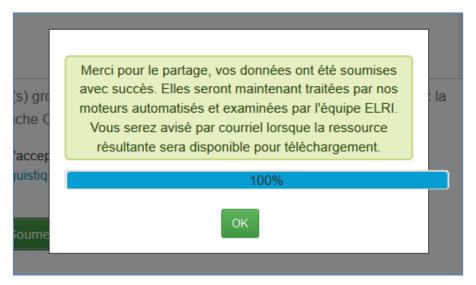


Figure 10 Partage de ressources réussi

À partir de ce moment, les fichiers chargés seront traités par les moteurs d'ELRI et les résultats ainsi que les informations fournies seront examinés par l'équipe d'ELRI en charge de la SRN. Une fois la ressource validée par l'équipe dédiée d'ELRI, qui s'assure que les résultats du traitement et les informations légales sont valides, la ressource sera publiée. L'utilisateur qui a fourni la ressource sera averti par courriel de sa publication, comme indiqué à la Figure 11.

```
Cher/Chère Utilisateur Dupont,

Votre ressource Corpus français-anglais de rapports annuels a été publiée sur la Station Relais Nationale ELRI. Vous pouvez maintenant vous connecter à notre site web <a href="https://elri.elda.org">https://elri.elda.org</a> et télécharger votre ressource préparée via la page Parcourir.

Merci d'utiliser les services ELRI, n'hésitez pas à nous contacter pour toute demande d'information.

Bien cordialement,

L'équipe ELRI
```

Figure 11 Courriel de notification pour la publication d'une ressource

Une fois publiée, la ressource sera accessible à partir du catalogue, à la fois pour l'utilisateur qui l'a déposé et pour tous les membres du ou des groupes sélectionnés, comme décrit à la section 5.2.

5.2 Téléchargement de ressources linguistiques

Pour accéder aux ressources linguistiques, les utilisateurs doivent cliquer sur l'option **Parcourir** de la barre de menu, qui ouvre la page du catalogue des ressources, comme indiqué dans la Figure 12.



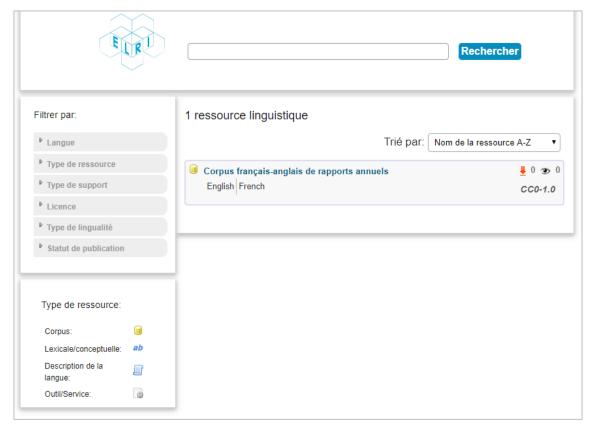


Figure 12 Page du catalogue des ressources

À partir de cette page, les utilisateurs peuvent rechercher des ressources spécifiques via la barre de recherche situé en haut de la page, visualiser les ressources disponibles page par page ou filtrer les ressources grâce aux options du panneau dédié, situé à gauche de la page.

Pour accéder à une ressource spécifique, les utilisateurs doivent simplement cliquer sur le nom de la ressource, ce qui ouvre une nouvelle page contenant des informations spécifiques sur la ressource, comme indiqué dans l'exemple de la Figure 13.



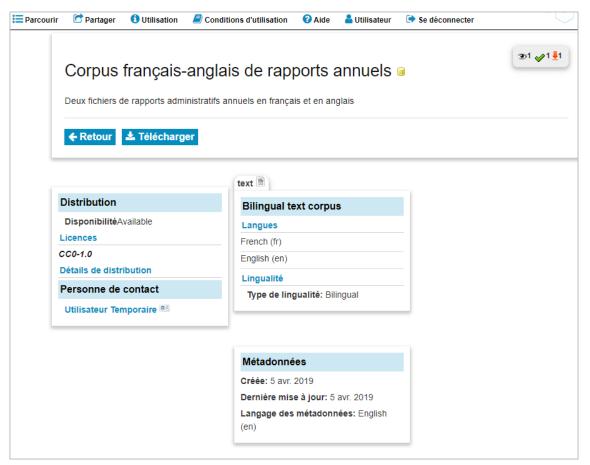


Figure 13 Exemple d'informations sur une ressource

Pour télécharger la ressource, les utilisateurs doivent cliquer sur le bouton **Télécharger**. Cela ouvre la page illustrée à la Figure 14, où l'utilisateur doit accepter les termes de la licence associés à cette ressource spécifique.





Figure 14 Page de la licence

Une fois la licence acceptée en cochant la case correspondante, le téléchargement peut avoir lieu et un fichier compressé dénommé *archive.zip* sera téléchargé. Ce fichier contient la ressource traitée, par exemple une mémoire de traduction au format TMX tel que présentée à la Figure 15.



Figure 15 Ressource téléchargée

6 Contacter l'équipe de la SRN

Pour toute question ou demande, l'équipe d'ELRI peut être contactée via la page d'Aide accessible depuis la barre de menu et en envoyant un courriel à l'adresse fournie, comme l'indique la Figure 16.



Coordonnées

- Pour toute question sur ce projet, merci d'utiliser la liste de contacts ELRI.
- Pour une assistance technique sur cette Station Relais Nationale, vous pouvez nous contacter à elri@elda.org.

Figure 16 Informations de contact