
Charte de l'Equipex ORTOLANG

1	Préambule	2
2	Engagements mutuels d'Ortolang, des contributeurs et des utilisateurs.....	2
3	Fonctionnalités et engagements d'Ortolang.....	3
4	Engagements des contributeurs.....	4
5	Engagements des utilisateurs.....	6
6	ANNEXE 1 : Fonctionnalités contenues de l'EquipEx ORTOLANG.....	7
6.1	Fonctionnalités du site www.ortolang.fr offerte sans identification	7
6.2	Fonctionnalités supplémentaires de la plateforme www.ortolang.fr après identification.....	7
7	ANNEXE 2 : Liste des types d'identification d'utilisateurs proposés.....	8

1 Préambule

ORTOLANG est un équipement d'excellence validé dans le cadre des investissements d'avenir initiés par le gouvernement français. Son but est de proposer une infrastructure en réseau¹ offrant un réservoir de ressources (corpus, lexiques, dictionnaires, etc.) et d'outils sur la langue et son traitement.

Ce réservoir doit être disponible et être alimenté le plus largement possible pour permettre :

- une mutualisation efficace des ressources de recherche sur le langage,
- une exploitation aisée par les scientifiques, les enseignants ou le grand public.

Les fonctionnalités et le contenu de la version 1 de l'EquipEx ORTOLANG et du site Internet www.ortolang.fr sont décrits par ailleurs (cf. annexe 1 de cette charte).

Le fonctionnement et la réussite d'un tel équipement dépendent autant de l'équipement que de ses usagers. C'est pour cela que cette charte d'utilisation de l'EquipEx ORTOLANG a été rédigée : son respect permet une utilisation optimale des moyens mis à la disposition des usagers d'ORTOLANG.

Dans le texte ci-dessous, on utilisera les termes suivants :

- **ORTOLANG** désigne les institutions et les personnels chargés de la gestion de la plateforme et du site Internet www.ortolang.fr, de la sécurisation et de la diffusion des ressources, de la création ou de la diffusion des logiciels et des outils figurant sur le site ;
- les **contributeurs** désignent les personnes ou les institutions qui fournissent des ressources informatisées qui alimentent le réservoir que représente ORTOLANG ;
- les **utilisateurs** désignent les personnes ou les institutions qui se servent des ressources disponibles sur ORTOLANG, que ce soit à des fins de recherche, d'enseignement ou pour tout autre usage, que ce soit par téléchargement partiel ou intégral, ou par simple consultation.

2 Engagements mutuels d'Ortolang, des contributeurs et des utilisateurs

L'EquipEx ORTOLANG est conscient que les contributeurs et les utilisateurs peuvent avoir des besoins et des contraintes très variés. C'est pour cela qu'il existe une large gamme de configurations concernant la protection des ressources et les conditions de leur utilisation.

Les contributeurs sont conscients que le dépôt de ressources sur une plateforme publique de gestion et d'archivage présente un coût important qui engage le contributeur à une diffusion des ressources.

¹ On trouvera à l'adresse www.ortolang.fr/#/information/partners la liste des partenaires porteurs de ce projet.

Les utilisateurs des ressources sont conscients que les investissements produits par l'ensemble des acteurs de la chaîne de création et de diffusion des corpus doivent être rendus visibles et de ce fait justifiés auprès des institutions qui financent les projets. Il faut également valoriser le travail de recherche et **respecter la propriété intellectuelle des contributeurs**. Pour cela, il faut **citer** les sources de ressources utilisées et faire largement connaître l'usage qui est fait de ces ressources.

L'application de ces principes permet une valorisation mutuelle de l'ensemble des moyens mis en œuvre par les acteurs de la recherche en linguistique, l'État, principal partenaire financier d'ORTOLANG, les contributeurs, et les utilisateurs.

3 Fonctionnalités et engagements d'Ortolang

Le dépôt de ressources est totalement gratuit, dans la limite des volumes qui sont à l'échelle des capacités d'ORTOLANG². Les ressources déposées peuvent couvrir toutes les sciences du langage, de l'écrit à l'oral et au multimodal.

ORTOLANG s'engage à réaliser une conservation sécurisée à long terme des dépôts numériques, en utilisant les moyens de l'informatique et du stockage sécurisé de ressources. La durée du stockage est au minimum celle de l'existence d'ORTOLANG, existence garantie par les institutions qui sont représentées dans ORTOLANG¹ (notamment Universités et CNRS).

Le dépôt de ressources est libre pour toute source de ressources provenant de laboratoires de recherche français³ et portant sur toute langue, ou pour toute source de ressources portant sur les langues de France,⁴ quelle que soit leur origine. Pour les sources d'autres origines, un accord ponctuel est possible sur la base d'une négociation entre les parties. Ce type d'accord peut amener à une participation financière.

ORTOLANG reçoit, sauvegarde et diffuse toutes les ressources qui concernent les sciences du langage. Les formats de ressources pouvant être traités ainsi ne sont pas limités que par les capacités techniques d'ORTOLANG.

Un certain nombre de formats de données bénéficient de traitements supplémentaires spécifiques de la linguistique. Il s'agit en particulier de corpus de langage fournis dans des formats connus et utilisés par la communauté scientifique (cette liste des formats connus peut s'enrichir dans le temps). Ces formats bénéficient d'outils comme par exemple l'indexation plein texte, le traitement automatique syntaxique, la visualisation en direct, etc. Des traitements automatiques sont proposés en priorité pour le français pour le français.

Les ressources qui sont dans des formats compatibles avec les exigences de l'archivage à long terme par les services des Archives nationales seront proposées à l'archivage

² Cette capacité évolue d'année en année. Début 2014 la capacité d'ORTOLANG est de 40To utiles de disque partagé utiles et de 312,5 To d'espace de sauvegarde.

³ C'est-à-dire dont le financement relève de l'Etat français.

⁴ La caractérisation « langue de France » relève des compétences de la DGLFLF.

auprès des institutions spécialisées au travers de la solution mise en place par la TGIR Huma-Num. L'archivage effectif par ces institutions est sous la seule responsabilité de ces institutions, ORTOLANG ne jouant que le rôle de service versant dans ce processus. Lorsque les archivages seront réalisés, une trace de ce dépôt sera incluse dans les métadonnées des ressources.

ORTOLANG s'engage à respecter les contraintes de sécurité des données définies par les contributeurs (voir ci-dessous). Pour cela, ORTOLANG s'engage à contrôler la modification des données par les contributeurs et leur exploitation par les utilisateurs.

Les ressources déposées sont identifiées de manière unique à l'aide d'un identifiant dit pérenne. Ceci permet de diffuser une référence à des ressources qui pourront être retrouvées à tout moment y compris de nombreuses années après le dépôt. L'identification des ressources de manière unique impose l'utilisation d'un système de versionnage. Les ressources déposées sont donc associées à un numéro de version unique qui fait partie de l'identifiant pérenne. La modification des ressources de manière mineure (correction d'erreurs par exemple) est possible sans changement de version. Une modification majeure peut amener à la création d'un nouveau numéro de version. Pour des raisons de coût, une nouvelle version doit présenter au moins 10% de données nouvelles.

À terme, dans la version finale de l'Equipex ORTOLANG (fin 2016), grâce à des outils de supervision et de statistiques ORTOLANG s'engage à rendre accessible aux contributeurs des statistiques d'accès et de téléchargement de leurs ressources.

4 Engagements des contributeurs

Les contributions à ORTOLANG ne peuvent être réalisées que par des personnes identifiées. La création d'un identifiant s'accompagnera de la création d'un espace de travail permettant le dépôt, la préparation et la mise en forme des contributions.

Le dépôt de ressources sur ORTOLANG implique d'avoir une connaissance exacte des droits d'utilisation des ressources, et, dans le cas où le contributeur n'est pas le propriétaire ou l'ayant droit juridique des ressources, avoir toutes autorisations de dépôt de la part du propriétaire ou de l'ayant droit juridique.

Ainsi le dépôt de ressources implique :

- de pouvoir prendre toutes décisions concernant l'utilisation et la diffusion du corpus (propriété intellectuelle en particulier) ;
- de disposer de toutes les informations concernant les sources des corpus et le consentement des personnes enregistrées ou filmées.

ORTOLANG étant un service public financé par l'État, il est nécessaire pour déposer des ressources d'accorder un droit d'utilisation de celles-ci qui comprenne au minimum celui de leur utilisation libre dans le cadre de la recherche scientifique publique soit par téléchargement soit au travers un outil spécifique de visualisation ou d'exploitation.

- Les droits supplémentaires sont largement encouragés et il est possible d'appliquer des droits différents pour des sous-ensembles des ressources déposées⁵. Les ressources déposées ne peuvent donner lieu à rétribution financière.
- Il est possible de déposer des ressources dont l'ouverture à l'utilisation par d'autres personnes que les contributeurs est fixée à une date ultérieure (système de quarantaine), soit pour des raisons d'exploitation des ressources et de retombées des travaux, soit pour des raisons juridiques.
- La durée des périodes de quarantaine liées à des contraintes juridiques ne dépend que de la loi française et n'est pas du ressort d'ORTOLANG. La durée des périodes de quarantaine liées à des droits intellectuels ou financiers est limitée à la durée des contraintes (bourse de thèse, financement ANR, etc.) plus une période maximale de deux ans. Au-delà de cette période, les droits d'utilisation fixés s'appliquent.

Dans tous les cas le dépôt d'une ressource sur ORTOLANG devra être accompagné d'un jeu minimum de métadonnées descriptives au format Dublin Core, et administratives spécifiques à ORTOLANG. Pour aider les utilisateurs à construire ces métadonnées, un éditeur interactif de ces métadonnées est proposé par ORTOLANG dans l'espace de travail lors du dépôt d'une ressource. Il permet de préciser en particulier des renseignements généraux de type descriptif de la ressource, les droits y afférant et les divers contributeurs à cette ressource. Le dépôt d'une ressource vaut acceptation de diffusion entièrement libre de ses métadonnées administratives et descriptives.

Le dépôt de ressources peut s'accompagner de consignes d'utilisation, notamment sur la manière de citer les ressources utilisées. La politique d'ORTOLANG est d'exiger que soit citée, a minima, la source des ressources, c'est-à-dire leur identifiant pérenne. La demande de citation de travaux, références scientifiques, indications de propriété intellectuelle est possible, dans la limite de trois citations par corpus. Le respect de ces citations fait partie de la charte d'usage des utilisateurs et doit être respecté par tous les acteurs du domaine.

Le dépôt de ressources par des étudiants ou thésards ne relevant pas d'un établissement scientifique public est possible et souhaitable, et doit se faire sous la responsabilité du laboratoire dans lequel ils réalisent leurs travaux.

Le dépôt de ressources incomplètes (ressources en cours de numérisation, ressources audiovisuelles non encore transcrites) est possible et souhaitable (en particulier pour des raisons de sécurité des ressources) à condition que les droits de propriété et d'utilisation des ressources soient clairement définis et décrits. Ce type d'usage permet la mise en place d'un système de quarantaine. Toutefois, un dépôt partiel et non finalisé de données sera rendu public et disponible à la recherche à la fin de la période de quarantaine, quel que soit le statut des données.

Pour des raisons de coût de stockage et de coût de fonctionnement, les ressources de type média (audio, vidéo, données physiologiques) ne peuvent être déposées que si elles

⁵ Cf. annexe 2 : liste des types d'identification d'utilisateurs proposés sur la plateforme.

sont associées à des annotations linguistiques. Si les ressources sont en cours de création (par exemple dans le cadre d'un projet de recherche), des données média peuvent être déposées par anticipation sans que les annotations soient déjà disponibles à condition que les droits d'utilisation soient déjà clairement définis.

Les ressources déposées ne sont modifiables que par le déposant original. La modification du déposant ou le transfert de droit s'accompagne de la création d'un nouveau numéro de version.

Les ressources déposées par un contributeur sans en posséder les droits pourront être supprimées d'ORTOLANG, notamment en cas de recours juridique ou de conflit de droit.

5 Engagements des utilisateurs

Les utilisateurs doivent s'identifier sur le site pour toute utilisation autre que celles des ressources libres de tout droit. Cette contrainte permet de contrôler et d'appliquer les droits définis par les contributeurs.

L'utilisation de ressources à des fins scientifiques ou personnelles doit respecter les contraintes fixées par les contributeurs. Elle doit également respecter les usages (voir la charte « Ethique et Big Data : <http://wiki.ethique-big-data.org/chartes/CharteEthiqueBigDataV8.pdf>).

La diffusion, mais aussi la visualisation des ressources doivent respecter les droits originaux. Il n'est ainsi pas possible de diffuser publiquement des ressources qui ont des droits restreints.

Toute ressource utilisée doit être accompagnée de son identifiant pérenne ORTOLANG et les citations souhaitées par les contributeurs doivent être respectées. Les identifiants et modes de citations des ressources sont précisés dans leurs métadonnées. Tout utilisateur d'une ressource ORTOLANG est donc tenu de les faire apparaître dans ses citations bibliographiques, ses remerciements, ses notes de bas de page ou sa licence, selon les conditions d'utilisation des ressources.

6 ANNEXE 1 : Fonctionnalités contenues de l'EquipEx ORTOLANG

6.1 Fonctionnalités du site www.ortolang.fr offertes sans identification

Le site www.ortolang.fr, bilingue français anglais, offrira à tout utilisateur les fonctionnalités suivantes :

- a) Accueil : présentant les actualités et les nouveautés sur ORTOLANG ainsi que diverses informations sur le projet (Présentation, partenaires, feuille de route, lettres d'information, charte d'ORTOLANG, mentions légales, accès au site communautaire, etc.).
- b) Présentation avec fenêtre de recherche des ressources intégrées dans ORTOLANG :
 - Corpus
 - Lexiques
 - Outils
 - Projets intégrés

Cela permet, pour chaque ressource :

- l'accès à une page descriptive de la ressource,
- l'accès à une visualisation d'un extrait
- et le téléchargement dans le respect des contraintes de diffusion de la ressource

6.2 Fonctionnalités supplémentaires de la plateforme www.ortolang.fr après identification

Après identification l'utilisateur pourra accéder à chacun de ses espaces de travail qui, pour une ressource, lui permettront :

- a) De déposer les fichiers composant sa ressource
- b) D'accéder aux divers fichiers composant sa ressource
- c) D'éditer ses métadonnées de présentation
- d) De sauvegarder son espace de travail
- e) De suivre les processus de traitement en cours
- f) D'accéder à des outils applicables à sa ressource
- g) D'obtenir des prévisualisations avant publication
- h) De visualiser un historique et les membres ayant accès à cet espace de travail
- i) De soumettre sa ressource à publication

7 ANNEXE 2 : Liste des types d'identification d'utilisateurs proposés

ORTOLANG proposera 4 classes d'identification des utilisateurs potentiels d'une ressource :

- a) L'ensemble des utilisateurs potentiels sans aucune restriction
- b) Les membres de l'ESR français
- c) L'ensemble des utilisateurs inscrits sur la plateforme
- d) Un groupe spécifique d'utilisateurs préalablement identifiés sur la plateforme ORTOLANG.