Euskararen erakunde publikoa
Office public de la langue basque

Développement d'un traducteur neuronal en partenariat avec le Gouvernement de la Communauté autonome d'Euskadi : Convention de mise à disposition d'un corpus

Itzultzaile automatiko neuronal baten garapena Eusko Jaurlaritzarekin lankidetzan : Corpus bat esku emateko hitzarmena

Contexte et présentation du projet :

Dans le cadre du partenariat en matière de politique linguistique formalisé avec le Gouvernement de la Communauté Autonome d'Euskadi (CAE), le Vice-ministère à la politique linguistique du gouvernement de la CAE a proposé à l'OPLB de développer un traducteur automatique neuronal français/basque – basque/français, en constituant un corpus de textes bilingues pour entrainer les « modèles de traduction » du traducteur. Cette technologie imite le fonctionnement du cerveau humain en analysant et comparant des données de différentes langues, puis génère des modèles de traduction permettant de créer des outils de traduction automatique de grande qualité.

Au vu de l'intérêt de cet outil, l'Assemblée générale de l'OPLB du 17 décembre 2019 a décidé d'approuver le projet le partenariat avec le Gouvernement de la CAE pour la création d'un traducteur automatique français – basque / basque – français, d'engager 88.000 € pour permettre le financement de cette opération sur le budget 2019 de l'OPLB et d'autoriser le recrutement d'un chargé de mission pour une année.

L'OPLB a engagé la mise en œuvre du projet et a constitué un corpus de 6.000 documents bilingues. Ces documents ne sont pas protégés par le droit de la propriété intellectuelle, ils ne comportent pas de données personnelles. A ce stade du projet, il convient de les transmettre au Gouvernement de la CAE. A cette fin, la convention annexée au présent rapport a été rédigée avec l'appui du Bureau de la Propriété Intellectuelle du Ministère de la Culture et de la Communication du Gouvernement de la CAE, qui définit les modalités de mise à disposition de ce corpus pour améliorer et entraîner le traducteur neuronal.

Testuingurua eta proiektuaren aurkezpena:

Hizkuntza politika alorrean Euskal Autonomia Erkidegoarekin (EAE) landu lankidetza lankidetzaren karietara, EAEko Gobernuko Hizkuntza Politikako sailburu ordetzak EEPri proposatu dio frantses/euskara, euskara/frantsesa itzultzaile automatiko neuronal bat lankidetzan garatzea, "itzulpen modeloak" garatzeko beharrezkoa den itzulpen corpus bat osatuz. Teknologia horrek giza garunaren funtzionamendua imitatzen du, hainbat hizkuntzatako datuak aztertuz eta konparatuz, eta kalitate handiko itzulpen automatikoko tresnak egiteko itzulpen-ereduak sortzen ditu.

Tresnaren garrantzia ikusirik, 2019-ko Abenduaren 17-ko EEPren Biltzar Nagusiak erabaki zuen EAErekin frantsesa – euskara / euskara – frantsesa itzultzaile neuronala garatzeko partaidetza proiektua onartzea; 2019ko aitzinkontuan 88.000 €-ko finantzamendua onartzea, eragiketa hori bideratzeko; behin-behineko lanpostu bat sortzea corpusaren sortze lanetan aritzeko.

EEPek proiektua abiatu du eta 6000 dokumentu elebiduneko corpusa osatu du. Dokumentu horiek jabetza intelektualaren babesetik kanpo dira eta ez dute datu pertsonalik. Proiektuaren urrats honetan, dokumentuak EAE-ko Gobernuaren esku jartzeko beharra agertzen da. Hori dela eta, txosten honen gaia den hitzarmena, Kultura eta Komunikazioaren Ministerioko Jabetza Intelektualaren Bulegoaren laguntzarekin idatzi dena, itzultzailea hobetzeko eta trebatzeko beharrezkoa den corpusaren eskuratze baldintzak zehazten ditu.

Délibération :

Compte tenu de ces éléments, l'Assemblée générale de l'OPLB décide :

- d'adopter le projet de convention annexé au présent rapport ;
- d'autoriser le Président de l'OPLB à signer cette convention.

Adopté à l'unanimité.

Erabakia:

Elementu hauek kontuan hartuz, EEPko Biltzar Nagusiak erabakitzen du:

- eranskinean agertzen den hitzarmena onartzea;
- EEPren Presidenteari hitzarmena sinatzeko baimena ematea.

Aho batez onartua.

CONVENTION DE MISE A DISPOSITION DU CORPUS LINGUISTIQUE ENTRE L'OPLB ET LA COMMUNAUTE AUTONOME D'EUSKADI

CORPUS LINGUISTIKO ESKURATZE HITZARMENA EEP ETA EUSKO JAURLARITZAREN ARTEAN

Entre les soussignés :

D'une part, **Miren Dobaran Urrutia**, Vice-ministre de la Culture et la Politique linguistique, représentant le Gouvernement de la Communauté Autonome d'Euskadi (ci-après désigné « Gouvernement de la CAE »).

Et

D'autre part, **Antton Curutcharry**, Président de l'Office Public de la Langue Basque (ci-après désigné « OPLB »).

Il est convenu et arrêté ce qui suit :

PRÉAMBULE

L'intégration du basque dans les Technologies de l'Information et de la Communication, plus largement dans l'environnement digital, est une ambition et un objectif constant du Gouvernement de la CAE.

Les traducteurs automatiques ont connu un bond technologique. Ils fonctionnent désormais grâce à la technologie des « réseaux neuronaux » de l'intelligence artificielle. Cette technologie utilise des correspondances de traduction pour entraîner des modèles servant de base à l'application du traducteur neuronal.

Le Gouvernement de la CAE a initié en 2007 un programme visant à développer un traducteur automatique neuronal.

Honako hauek izenpeturik:

Batetik, **Miren Dobaran Urrutia** andreak, Eusko Jaurlaritzako Hizkuntza Politikarako sailburuordeak, Eusko Jaurlaritzaren ordezkari eta haren izenean.

Eta

Bestetik, **Antton Curutcharry** jaunak, Euskararen Erakunde Publikoaren (ondotik «EEP» deitua) Presidente eta haren izenean.

Bi alderdiek honako hau onartu dute:

SAR HITZA

Hizkuntzen teknologiak agertu zirenetik, Eusko Jaurlaritzaren ardura eta helburua izan da, beti, euskara Informazioaren eta Komunikazioaren Teknologietan eta, oro har, Ingurune digitalean integratzea.

Itzulpen automatikoaren alorrean, sekulako jauzia gertatu da. Itzulpen automatikoak, gaur egun, erabiltzen duen teknologia adimen artifiziala da, hain zuzen «sare neuronaletan» oinarritzen dena. Teknologia honek itzulpen korrespondentziak baliatzen ditu aplikazioaren oinarri diren modeloak trebatzeko.

Itzulpen automatikoari dagokionez, 2007an hasi zen Eusko Jaurlaritza lanean itzultzaile automatikoa garatzeko lehenengo urratsak ematen.

En 2009, un premier traducteur automatique a permis de réaliser des traductions du castillan vers le basque. Ce traducteur a été mis en ligne en 2012. Depuis 2019 le traducteur automatique neuronal permet aux usagers les traductions suivantes : euskara ↔ castillan, français ↔ euskara et anglais ↔ euskara (www.euskadi.eus/itzuli).

L'OPLB est un Groupement d'Intérêt Public constitué par l'État, la Région Nouvelle-Aquitaine, le Département des Pyrénées-Atlantiques et la Communauté d'Agglomération Pays Basque. Sa mission est de définir et mettre en œuvre une politique publique en faveur de la langue basque.

L'OPLB possède, dans son domaine d'activité, des textes de traduction bilingues permettant d'accroitre les bases de traduction.

Selon le Décret 18/2020 du 6 septembre signé par le Président du Gouvernement de la CAE, la mise en œuvre de la politique linguistique et la promotion du basque sont des missions et des compétences relevant du Ministère de la Culture et de la Politique Linguistique.

Dans le cadre du partenariat en matière de politique linguistique formalisé avec le Gouvernement de la Communauté Autonome d'Euskadi (CAE), le Vice-ministère à la politique linguistique du gouvernement de la CAE a proposé à l'OPLB de développer un traducteur automatique neuronal français/basque – basque/français, en constituant un corpus de textes bilingues pour entrainer les « modèles de traduction » du traducteur.

Le 19 décembre 2019, l'Assemblée générale de l'OPLB a décidé d'approuver le projet de partenariat avec le Gouvernement de la communauté autonome d'Euskadi pour la création d'un traducteur automatique français/basque et basque/français et d'engager sur le budget 2019 de l'OPLB une enveloppe financière de 88.000 € pour permettre le financement de l'opération.

L'OPLB a engagé la mise en œuvre du projet et a constitué un corpus de 6.000 documents bilingues. Ces documents ne sont pas protégés par le droit de la propriété intellectuelle, ils ne comportent pas de données personnelles. A ce stade du projet, il convient de les transmettre au Gouvernement de la CAE.

2009an hasi zen garatzen gaztelania-euskara itzulpen-norabidea. 2012an jarri zen herritarren eskura itzultzaile hori. 2019ko irailaz geroztik itzultzaile automatiko neuronala herritar guztion eskura dago itzulpen-norabide hauetan: euskara↔ gaztelania, frantsesa↔euskara eta ingelesa↔euskara (www.euskadi.eus/itzuli).

EEP Estatuak, Akitaniako Eskualdeak, Pirinio Atlantikoetako Departamenduak eta Euskal Elkargoak osatzen duten interes publikoko elkargoa da. Euskararen aldeko politika publiko adostua definitzea eta obratzea du eginkizuna. Eginkizun horri erantzuteko gisan, Euskararen erakunde publikoak hizkuntza politika proiektu bat definitu du.

EEPk itzulpen oinarriak garatzeko baliagarriak diren testu elebidunak ditu bere eskumeneko eremuan.

Eusko Jaurlaritzako Lehendakariaren irailaren 6ko 18/2020 Dekretuaren arabera, Kultura eta Hizkuntza Politika Sailari, besteak beste, honako egiteko eta jardun-arlo hauek dagozkio: hizkuntza-politika egitea eta euskara sustatzea.

Hizkuntza politika alorrean Euskal Autonomia Erkidegoarekin (EAE) landu lankidetza lankidetzaren karietara, EAEko Gobernuko Hizkuntza Politikako sailburu ordetzak EEPri proposatu dio frantses/euskara, euskara/frantsesa itzultzaile automatiko neuronal bat lankidetzan garatzea, "itzulpen modeloak" garatzeko beharrezkoa den itzulpen corpus bat osatuz.

EEPko 2019ko abenduaren 19ko Biltzar Nagusiak, Eusko Jaurlaritzarekin, euskara/frantsesa eta frantsesa/euskara itzultzaile automatiko bat garatzeko lankidetza proiektua onartu zuen eta ekintza honen finantzatzeko beharrezkoa zen 88.000 €-ko aurrekontua onartu zuen.

EEPek proiektua abiatu du eta 6000 dokumentu elebiduneko corpusa osatu du. Dokumentu horiek jabetza intelektualaren babesetik kanpo dira eta ez dute datu pertsonalik. Proiektuaren urrats honetan, dokumentuak EAE-ko Gobernuaren esku jartzeko beharra agertzen da.

Article 1 : Objet de la convention

La convention a pour objet de définir les modalités de mise à disposition du corpus de traduction de l'OPLB au Gouvernement de la CAE, corpus nécessaire pour améliorer et entraîner le traducteur neuronal.

Article 2 : Durée de validité de la convention

La convention a une durée de validité de 4 ans, à compter de la date de signature par le dernier signataire, elle pourra être renouvelée pour 4 années si les parties en expriment la volonté.

Article 3 : Engagements du Gouvernement de la CAE

Le Gouvernement de la CAE s'engage à ce que les documents du corpus mis à sa disposition soient utilisés pour améliorer et entraîner le traducteur automatique neuronal tel que décrit au préambule de la présente convention.

Le Gouvernement de la CAE mentionnera l'OPLB sur le site web du traducteur automatique neuronal.

Le Gouvernement de la CAE donnera accès au traducteur neuronal à l'OPLB sur son site web, à travers une application.

Article 4 : Engagements de l'OPLB

L'OPLB met à disposition du Gouvernement de la CAE le corpus plurilingue collecté pour l'entraînement du traducteur neuronal.

Le corpus sera constitué de documents appartenant à l'OPLB et de documents collectés auprès d'opérateurs du territoire français dans le respect du droit de la propriété intellectuelle et de la protection des données personnelles.

L'OPLB fera figurer sur son site internet, le Gouvernement de la CAE comme organisme ayant développé le projet de traducteur automatique.

Article 5 : Propriété du corpus de traduction - propriété intellectuelle

L'OPLB transmet des documents non protégés par le droit de propriété intellectuelle et sans données personnelles. Pour rappel, les documents courants de gestion, administratifs et officiels ne sont pas soumis au droit de propriété intellectuelle.

1. artikulua: Hitzarmenaren objektua

Hitzarmen honek, EEPren eta Eusko Jaurlaritzaren arteko, itzultzaile neuronala hobetzeko eta trebatzeko beharrezkoa den corpusaren lagatze baldintzak zehazten ditu.

2. artikulua: Hitzarmenaren indarraldia

Hitzarmena 4 urtez izango da indarrean, azken izenpetzailearen sinaduraren egunetik hasita, eta beste 4 urtez luzatu daiteke bi aldeek horretarako berariazko borondatea adierazten badute.

3. artikulua: Eusko Jaurlaritzak bere gain hartzen dituen betebeharrak

Eusko Jaurlaritzak adierazten du bere esku jarriko den corpus-eko dokumentuak, itzultzaile neuronala hobetzeko eta trebatzeko erabiliak izanen direla, hitzarmenaren aintzin solasan zehaztu bezala.

Eusko Jaurlaritzak itzultzaile automatiko neuronalaren webgunean EEPren aipamena egingo du.

Eusko Jaurlaritzak API baten bidez sarbidea emango dio EEPri itzultzaile neuronalera, bere web-orrialdean.

4. artikulua: EPPk bere gain hartzen dituen betebeharrak

EEPk Eusko Jaurlaritzaren esku jarriko du itzultzaile neuronala garatzeko bildu duen corpus-a.

Corpusak EEPren dokumentuak eta Frantziako lurraldeko eragileen bidez bildutakoak izango ditu, betiere, jabetza intelektualaren eta datu pertsonalen babesaren eskubidea errespetatuz.

EEPk, Eusko Jaurlaritzaren aipamena eginen du, itzultzailea eskuragarri izanen den web orrialdean, proiektua garatu duen erakunde bezala.

5. artikulua: Corpus-aren jabetza intelektuala-jabetza intelektuala

EEPk ez ditu jabetza intelektualak babestutako edo datu pertsonalak dituzten dokumentuak Eusko Jaurlaritzaren esku jarriko. Oroitgarri gisara, kudeaketa dokumentu arruntak, administrazio dokumentuak eta dokumentu ofizialak ez dira jabetza intelektualaren zuzenbidearen menpekoak.

Article 6 : Non-respect des obligations et engagements souscrits par chacune des parties

Le non-respect des engagements des signataires entraîne la résiliation de la convention.

6. artikulua: Alderdietako bakoitzak bere gain hartutako betebehar eta konpromisoen ez betetzeari aplikatu beharreko ondorioak

Alderdiek hartutako betebeharrak eta konpromisoak beteko ez balira, hitzarmena bertan behera geratuko litzateke.

Article 7 : Régime de modification de la convention

En l'absence de disposition expresse, la modification du contenu de la convention nécessite l'accord des parties.

7. artikulua: Hitzarmena aldatzeko araubidea.

Arau espresurik ez badago, sinatzaileen ahobatezko erabakia beharko da hitzarmenaren edukia aldatzeko.

Article 8 : Responsabilité de la gestion, du suivi, de la surveillance et du contrôle de l'exécution de la convention

La Direction de Promotion de la langue basque du Gouvernement de la CAE sera responsable du suivi de l'usage du corpus de textes transmis par l'OPLB. Elle informera l'OPLB de l'avancée du projet.

8. artikulua: Hitzarmenaren kudeaketaren, jarraipenaren, zaintzaren eta betetzearen ardura.

Eusko Jaurlaritzaren Euskara Sustatzeko Zuzendaritza izango da EEPk eskuragarri jarriko duen testu corpus-aren erabileraren arduraduna. EEP Proiektuaren aintzinamenduaz informatuko du.

Article 9 : Loi applicable – Attribution de compétence

La présente convention est soumise à la loi française.

Tout litige résultant de l'interprétation ou de l'exécution de la présente convention sera résolu par le tribunal administratif de Pau.

Afin que les éléments convenus apparaissent et soient exécutés tels que convenus, les partenaires signent chacune des pages des deux exemplaires du document.

9. artikulua: Lege egokigarria - Gaitasun ahalmena

Hitzarmen honek Frantziako legediari erantzuten dio.

Hitzarmen hau interpretatzean edo betearaztean sortzen den edozein auzi Paueko Epaitegi Administratiboak ebatziko dute kasuen arabera.

Hitzartutakoa hala ager dadin, eta esan bezala betetzeko, partaideek honako dokumentu honen hiru ale izenpetzen dituzte orrialde guztietan.

A Bayonne, le 2021,

Baionan, 2021eko …….aren ……..a,

Le Président de l'Office Public de la Langue Basque
Euskararen Erakunde Publikoko Presidentea

Antton Curutcharry

La Vice-ministre de la Culture et de la Politique Linguistique du Gouvernement Basque
Eusko Jaurlaritzako Kultura eta Hizkuntza Politikako sailburuordea

Miren Dobaran Urrutia