



Exploitation d'une base lexicale dans le cadre de la conception de l'ENPA Innovalangues

Mathieu Mangeot, Valérie Bellynck, Emmanuelle Eggers, Mathieu Loiseau,
Yoann Goudin

► To cite this version:

Mathieu Mangeot, Valérie Bellynck, Emmanuelle Eggers, Mathieu Loiseau, Yoann Goudin. Exploitation d'une base lexicale dans le cadre de la conception de l'ENPA Innovalangues. Ivan Smilauer; Jovan Kostov. Enseignement des Langues et TAL, Jul 2016, Paris, France. Actes de la conférence conjointe JEP-TALN-RECITAL 2016, 09, pp.48-64, 2016, ELTAL. <<https://jep-taln2016.limsi.fr/actes/Actes>

HAL Id: hal-01382457

<https://hal.archives-ouvertes.fr/hal-01382457>

Submitted on 17 Oct 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Exploitation d'une base lexicale dans le cadre de la conception de l'ENPA Innovalangues

Mathieu Mangeot^{1,4} Valérie Bellynck¹ Emmanuelle Eggers⁴

Mathieu Loiseau^{2,4} Yoann Goudin^{3,4}

(1) LIG, Bâtiment IMAG, 700 avenue Centrale - Domaine Universitaire, 38400 Saint-Martin-d'Hères, France

(2) LIDILEM, Université Grenoble Alpes, CS40700, 38058 Grenoble cedex 9

(3) CERLOM, INALCO, 2 rue de Lille, 75007 Paris

(4) IDEFI Innovalangues, Maison des Langues et des Cultures, 1141 avenue Centrale - Domaine Universitaire, 38400 Saint-Martin-d'Hères, France

valerie.bellynck@imag.fr, emmanuelle.eggers@univ-grenoble-alpes.fr,

mathieu.loiseau@univ-grenoble-alpes.fr, mathieu.mangeot@imag.fr,

yoann.goudin@univ-grenoble-alpes.fr

RÉSUMÉ

Le projet Innovalangues a pour but la conception d'un environnement numérique personnalisé d'apprentissage des langues (ENPA). Dans ce cadre, plusieurs modules tels des jeux sérieux ou des générateurs d'exercices ont besoin d'une base lexicale. Il est intéressant également de donner la possibilité aux apprenants de gérer leurs propres lexiques. Pour éviter que chacun ne développe sa propre base lexicale, il apparaît rapidement indispensable de développer une seule base lexicale commune qui puisse servir aux modules (machines) comme aux apprenants (humains). Après une analyse des besoins autour de scénarios d'utilisation, nous proposons une architecture de base lexicale multilingue. Nous avons ensuite réalisé un prototype fonctionnel qui s'intègre à l'ENPA et permet à un apprenant de consulter des ressources lexicales existantes et de créer son propre lexique. Le prototype LexInnova utilise la plate-forme Jibiki de gestion de ressources lexicales hétérogènes à distance via son interface de programmation (API) REST.

ABSTRACT

Operating a lexical database in the framework of the Innovalangues language learning platform

The Innovalangues project aims to design a personalized digital environment for language learning. In this context, several modules such as serious games or exercise generators need a lexical database. It is also interesting to provide solutions for learners to manage their own lexicons. To prevent each one to develop its own lexical database, it quickly appears essential to develop one common lexical database that can be used for modules (machines) and learners (humans). After an analysis around usage scenarios, we set up a multilingual lexical database architecture. We then built a working prototype that integrates with the environment and allows a learner to look up existing lexical resources and create its own lexicon. The LexInnova prototype uses the Jibiki platform for managing heterogeneous lexical resources via its REST application programming interface.

MOTS-CLÉS : Innovalangues, base lexicale, dictionnaire, lexique, Jibiki, LexInnova

KEYWORDS : Innovalangues, lexical database, dictionary, lexicon, Jibiki, LexInnova

1 Introduction

L'un des principaux objectifs du projet IDEFI Innovalangues est de mettre à disposition un environnement numérique personnalisé d'apprentissage des langues (ENPA). Outre les parcours d'apprentissage, l'écosystème numérique proposé par Innovalangues fournira aux apprenants des outils pour soutenir l'apprentissage : des jeux, des générateurs d'exercices, un chat, un test de positionnement (test automatique adaptatif pour l'orientation de l'apprenant vers un niveau cible en fin de formation), etc.

Le projet est divisé en plusieurs lots (donnant lieu à des livrables du projet) dont l'objectif est de produire des outils répondant à des problématiques didactiques et liés ou intégrés à l'ENPA.

L'une des tâches fondamentales de l'apprentissage des langues est de travailler le lexique¹. Dès lors, il est rapidement apparu que la réalisation des divers lots (outils) conduisait au développement de d'autant de bases lexicales avec des contraintes spécifiques pour chacune, sans qu'il n'y ait de lien pratique entre eux. De plus, les apprenants peuvent souhaiter également créer leurs propres lexiques d'apprentissage. Cette situation a rapidement posé un problème de mise en œuvre informatique (opérationnalisation), ce qui a mené à la définition d'un sous projet transversal au projet Innovalangues (« chantier interlot »), que nous avons appelé LexInnova.

Nous présentons d'abord le projet Innovalangues et les besoins les plus emblématiques en termes de ressources lexicales. La partie suivante concerne le cahier des charges de l'outil que nous souhaitons construire ainsi que les différents scénarios d'utilisation des ressources lexicales. Ensuite, nous spécifions la base lexicale multilingue visée. Enfin, dans la dernière partie, nous détaillons le fonctionnement du prototype fonctionnel que nous avons implémenté afin de montrer les possibilités qu'offre l'intégration un tel outil.

2 Présentation du projet Innovalangues

Le projet IDEFI Innovalangues, initié en 2012, est piloté par le service LANSAD de l'Université Grenoble-Alpes. Il tire son origine des difficultés des étudiants à atteindre un niveau B2 du Cadre Européen Commun de Référence pour les Langues (CECRL) (CECRL, 2000) alors qu'il s'agit du niveau cible pour le baccalauréat² (Masperi et Quintin, 2014a : 62). À partir de ce constat, le projet a pour but de fournir aux établissements supérieurs des moyens (dispositifs, méthodes, contenus) pour porter le degré de maîtrise en langues des étudiants à un niveau B2 certifié. Pour y parvenir, Innovalangues s'est fixé trois objectifs principaux :

- « se doter d'un environnement (numérique) personnalisé d'apprentissage (ENPA), “open source”, [proposant des contenus sous licence libre], au service de la collaboration, particulièrement flexible et hautement adaptable aux différents contextes rencontrés sur le terrain de la formation » (Masperi et Quintin, 2014b : 8) ;
- proposer un espace dédié aux enseignants et ainsi « capitaliser les résultats de la recherche en didactique des langues » (Masperi et Quintin, 2014a : 72) par des actions de formation et / ou via l'outillage techno-pédagogique ;
- créer une communauté de pratiques pour « étendre la dynamique de recherche-action en langues entreprise au niveau local » (Masperi et Quintin, 2014a : 72).

Le projet se déroule principalement en deux phases. La première phase est orientée conception/implémentation. Elle est prise en charge au sein de « lots » impliquant chercheurs en didactique, enseignants, concepteurs, développeurs. Les lots sont les suivants :

¹ Voir par exemple (Luste-Chaa, 2009 : 2.4) pour une présentation historique de la place du lexique en FLE.

² http://www.education.gouv.fr/cid206/les-langues-vivantes-etrangeres.html#Au_lycée_général_technologique_et_professionnel

- SELF (Système d'Évaluation en Langues à visée Formative) vise à mettre en œuvre une méthodologie de création de tests de positionnement et à fournir la plate-forme permettant de les accueillir pour six langues (anglais, italien, mandarin, japonais, espagnol, français langue étrangère).
- Innovason regroupe les activités issues de :
 - THEMPPPO (Thématique Prosodie & Production Orale) qui visait à proposer des outils et des pratiques centrées sur le développement des compétences prosodiques de la production orale en langue étrangère.
 - et COCA (Compétence Orale : Conception et Assistance) qui se fixait pour objectif de développer des solutions technologiques pour assister l'enseignant dans la conception d'activités de compréhension de l'oral.
- PARCOURS est centré sur le développement d'un module de l'ENPA (plate-forme fondée sur Claroline Connect) destiné à accueillir les parcours et réalisations des autres lots, mis à disposition de l'ensemble des acteurs de la formation.
- GAMER (Gaming Application for Multilingual Educational Resources) a pour objectif premier de concevoir et développer des jeux pour l'enseignement/apprentissage des langues, mais aussi de proposer des formations de formateurs sur l'usage du jeu en classe de langues.

Dans une seconde phase, les « équipes langues » (anglais, espagnol, italien, japonais, mandarin) s'approprient les outils conceptuels et technologiques conçus par les lots pour développer des contenus de formation. Pour organiser les développements autant que pour laisser mûrir les réflexions, les changements de phases ne sont pas complètement synchronisés (la seconde phase du lot SELF ne commence pas en même temps que celle du lot GAMER) et les phases ne sont pas hermétiques : équipes « lots » et « langues » collaborent durant les deux phases. Ces collaborations peuvent faire émerger un besoin différent et un chantier est alors créé pour explorer ce besoin. LexInnova s'inscrit dans ce type de travaux : il s'agit d'un chantier émergent visant à proposer des solutions concernant des enjeux partagés par plusieurs lots. L'équipe responsable de ce chantier est constituée d'enseignants de langue, de didacticiens et d'informaticiens spécialistes du TAL, tous issus de différentes équipes, d'Innovalangues et d'ailleurs.

Au sein du projet Innovalangues, les attentes en termes d'outillage du lexique s'expriment différemment : pour certains, les besoins principaux portent sur le contenu informationnel (voir par exemple Innovason, pour qui la transcription phonétique des mots du lexique semble primordiale), alors que d'autres utilisateurs potentiels sont plus tournés vers les fonctionnalités offertes par le système. Par exemple, pour le chantier Kinéphones, l'une des demandes concernait « l'écriture en couleurs » : une représentation différente des transcriptions phonétiques alignées sur la graphie où chaque phonème est associé à une couleur sur le modèle affranchi des contraintes éditoriales matérielles des tableaux de mots de l'approche Silent Way de Caleb Gattegno dématérialisé par Kinéphones³.

3 Recueil des besoins

Après avoir recueilli les informations linguistiques pertinentes pour les collègues du projet par le biais d'une page du wiki interne d'Innovalangues, nous avons décidé, dans une perspective de « techniques » de conception centrées utilisateur (Bastien et Scapin, 2004 : 461), d'élaborer conjointement des scénarios d'utilisation visant à préciser les besoins fonctionnels. Ceux-ci s'appuient sur des cas pratiques et peuvent intégrer des éléments biographiques exemplifiant les interactions entre langue « maternelle » et langue « cible ».

³ http://kinephones.u-grenoble3.fr/#/en_gb/1/table

3.1 Espagnol

Dans l'exemple de scénario suivant, légèrement retravaillé pour que le lecteur puisse en comprendre les enjeux, nous tentons de montrer comment la structure des lexiques et les besoins fonctionnels peuvent être intégrés à un scénario.

Anne étudie l'espagnol débutant depuis deux mois en LANSAD. Son enseignant a choisi de faire créer un lexique de groupe à tour de rôle par des étudiants en tandem. Pour ce faire, cette semaine, Anne et son partenaire travaillent sur l'ENPA. Ils relisent les notes prises durant le cours à la lumière des ressources proposées par l'enseignant et choisissent les mots qu'ils veulent inclure dans le lexique de groupe (d'un simple clic, quand la ressource est intégrée à l'ENPA). Dans certains cas, ils ajoutent à la notice de l'entrée un commentaire qui a été fait pendant la classe (ex : « Attention, en espagnol, '*padres*' signifie non seulement '*pères*', mais aussi '*parents*', ex : "*Llamo a mis padres cada semana.*" ; (voir aussi *hermanos*, *tíos*) » dans l'article '*padre*').

Ils profitent également de ce travail à destination du groupe pour mettre à jour leur propre lexique et cliquent sur les mots qu'ils souhaitent rajouter dans leur lexique respectif (lexique personnel qui alimente le lexique de groupe et *vice-versa*). Anne souhaite par exemple s'approprier le mot « *tío* » (oncle). Elle clique sur ce mot, un menu déroulant s'ouvre. Elle peut alors consulter la ou les définitions ainsi que les exemples correspondant à ce mot (structure), puis importer ces données telles quelles ou en les modifiant. Anne décide par exemple de modifier l'exemple associé à « *tío* » pour son lexique personnel ; son oncle s'appelant Bernard, pour retenir « *tío* » elle veut l'associer à « Bernard ». Elle va donc choisir de mettre comme exemple personnel « *Mi tío se llama Bernard* » (besoin fonctionnel). Anne va également intégrer dans son lexique personnel des mots pris en notes durant le cours et durant l'écoute des enregistrements audio mis à disposition dans le cadre de sa formation. Anne, pour réviser le lexique de groupe et son lexique personnel, peut lancer des générateurs d'exercices (lien avec l'ENPA). La semaine suivante, un autre tandem se chargera du lexique de la semaine.

Eunice, bilingue français-portugais, est inscrite dans le même cours. Elle va consulter les mots nouvellement arrivés dans le lexique de groupe et faire des exercices à partir de ces mots (besoin fonctionnel). Elle choisira de rajouter une information spécifique dans son lexique personnel « Attention : accent sur le "i" de "tío" » (contrairement au portugais).

3.2 Mandarin

Les scénarios pour un lexique aussi distant que celui du mandarin ne diffèrent fondamentalement pas de ceux conçus pour les autres langues. En revanche, il existe un besoin impérieux que la prise en charge du lexique puisse au-delà — ou plus précisément en-deçà — de l'entrée dans le lexique, traiter un certain nombre d'éléments tels que les sinogrammes dont chaque entrée est constituée. En effet, en sus de la compétence lexicale au cœur de notre démarche, la compétence graphique est également traitée à travers des informations relatives à la combinatoire des sinogrammes, et à un niveau encore inférieur incluant les composants graphiques, les traits ou encore les différentes lectures sino-xéniques des sinogrammes dont sont composées les entrées du lexique.

Ainsi, Pierre, apprenant en mandarin, pourra accomplir les mêmes opérations que celles décrites plus haut. 家 *jiā* – « famille », « maison » – vu en classe, figurera parmi les entrées de son lexique personnel. Plus tard, au détour d'une ressource en compréhension écrite, 國家 *guójiā*, « pays » portera l'information complémentaire selon laquelle cette entrée est programmée au cours de la formation, le système étant capable d'intégrer le lexique des unités à venir – le *lexique cible* – incitant ainsi les apprenants à s'y intéresser au plus tôt.

L'apprenant pourra également confronter les sinogrammes de 國家 *guójiā* à d'autres lexiques compilés par des institutions, qui associent à un niveau donné une liste de sinogrammes choisis selon différents principes. En effet, le paradigme didactique dominant du mandarin langue étrangère établit le sinogramme comme l'unité didactique fondamentale (Bellassen, 2010) induisant une chronopédagogie qui programme l'enseignement-apprentissage des sinogrammes selon un double

critère de fréquence d'un sinogramme dans le corpus et de la haute combinabilité de ce même sinogramme dans le lexique contemporain. Le corollaire d'une telle pratique fige des listes de sinogrammes par niveau qui constituent le socle de l'industrie certificative. Si l'approche didactique retenue pour Innovalangues veut rompre avec ce paradigme, nous ne pouvons pas exclure ces listes que nous étiquetons dans le système comme le *lexique institutionnel* informant l'utilisateur des coordonnées de chaque sinogramme dans différents contextes institutionnels : programmes de langue vivante de l'Éducation nationale, niveaux du test certificatif de chinois HSK etc.

Plus spécifique aux langues sinogrammiques en se distinguant du paradigme évoqué ci-dessus, LexInnova viserait à offrir une vue sur le lexique qui indiquerait les entrées qui partagent un même sinogramme. Ainsi, avec notre exemple, 家 apparaît également dans les entrées 作家 *zuòjiā* « écrivain », 學家 *xuéjiā* « universitaire », 文學家 *wénxuéjiā* « spécialiste de littérature » etc. ou encore 家庭 *jiātíng* « famille », 家族 *jiāzú* « clan » etc. Cette vue sur le lexique permet ainsi de traiter les dimensions suivantes : la polysémie de 家, qui en plus de vouloir dire « famille » renvoie par ailleurs à la notion d'agent lorsque qu'il se situe en dernière position. Il est même indiqué que la position de 家 dans sa première combinatoire 國家 *guójiā* « pays » est une exception : en règle générale, et largement plus représentée dans le corpus, 家 renvoie à l'idée d'agent même si le paradigme communicationnel le présentera avant tout comme signifiant « famille ».

Au-delà de la mise en relation des différentes entrées contenant un même sinogramme, idéalement le système permettrait d'accéder aux informations suivantes : 家 est un sinogramme dont la structure est binaire verticale 冫, constitué des composants 宀 *miǎn* « le toit », et 豕 *shǐ* « le porc ». Il est même possible d'intégrer une notice grammatologique infirmant l'édifiante étymologie orientaliste selon laquelle, pour les « Chinois des temps anciens, la famille était représentée au moyen d'un cochon sous le toit » alors que depuis le IIe siècle EC, nous savons que 家 s'est d'abord noté 豕 au moyen du composant du 豕 « porc » et d'un autre composant phonétique – 段 *jiǎ* – encore quasi-homophone vingt siècles plus tard. A la faveur d'un processus orthographique inévitable sur une période historique aussi longue, le composant phonétique est tombé au profit d'un *composant discriminant* connu également sous le terme de « clé » : celui du toit 宀.

Enfin, le sinogramme est recontextualisé avec les sinogrammes qui partagent le même composant phonétique ou *phonophore* tel que 嫁 *jià* « prendre pour mari » ; mais également ses homophones parfaits ou au ton près tel que 加 *jiā* « ajouter » ou 價 *jià* « valeur, prix », 假 *jià* « vacances » ou 假 *jiǎ* « faux » – qui partage le même phonophore historique de 家 *jiā* – et les liens vers les entrées contenant ces sinogrammes. Finalement, 家 *jiā* est recontextualisé en intercompréhension avec les lectures sino-xéniques et d'autres langues sinitiques, *ka* – ou plus rarement *ke* – en sino-japonais, *ga* en sino-coréen, *gia* en sino-vietnamien, et même en français à travers l'ethnonyme *hakka* – 客家 *kèjiā* – désignant la communauté culturelle des vallées de l'arrière-pays des provinces méridionales du Fujian et du Guangdong, très présente dans la diaspora et majoritaire dans les confettis insulaires de l'empire français⁴.

Toutes ces fonctionnalités et vues sur le lexique permettent ainsi de sous-traiter un maximum d'informations à la plate-forme et ainsi consacrer le plus possible du cours en présentiel à l'interaction orale plutôt qu'aux explications de vocabulaire et autres digressions graphiques et culturelles qui rendent le mandarin une langue encore plus distante pour les apprenants. Par ailleurs, une telle prise en charge du lexique permet de modéliser le profil de l'apprenant pour la génération d'exercices ainsi que les jeux évoqués ci-dessus.

⁴ Le même changement phonétique d'une occlusive vélaire à une palatale est observable en français à travers l'emprunt Pékin au XVIIIe siècle devenu depuis Beijing...

3.3 Relations avec GAMER

Les deux exemples précédents montrent les liens entre l'ENPA et les lexiques dans deux langues. Le cas du mandarin illustre plus spécifiquement les enjeux liés à la forme écrite (les sinogrammes) des mots-forme, alors que le cas de l'espagnol, il s'agit de l'exploitation à partir des entrées lexicales (lemmes ou mots-vedettes). C'est ce type de lien avec l'ENPA qui est au cœur de la problématique du lexique dans le lot GAMER. Dans le cadre de ces travaux, plusieurs prototypes de jeux sont en cours de développement avec un point de vue différent sur le lexique. Par exemple, dans le cadre du jeu « Magic Word⁵ », dont les règles le placent dans la famille de jeux de lettres que l'on pourrait exemplifier par le Boggle (Loiseau, Zampa et Rebourgeon, 2015), il est nécessaire d'avoir une ressource lexicale qui définisse le champ des possibles (liste des formes autorisées). Ce jeu est focalisé sur ce que Portine appelle le pôle *accuracy*, c'est-à-dire sur le respect des schémas lexicaux et grammaticaux (Portine, 2013 : 162). Chaque langue prise en charge devra fournir une telle ressource contenant non-seulement « toutes » les formes de la langue, mais également les traits morphologiques associés et le lemme dont la forme est issue : certaines règles des futures versions du jeu en dépendent. Mais un autre besoin est la possibilité de lier les lemmes (associés aux formes trouvées dans le jeu) à l'activité du joueur dans les parties plus formelles de l'ENPA. Dans ce cas-là les générateurs d'exercices mentionnés dans le scénario espagnol pourraient être lancés à partir de cette liste de mots (dans ce contexte, le lexique du joueur, constitué automatiquement d'après les traces d'interaction ferait office de profil de l'apprenant).

Un autre exemple de jeu, radicalement différent dans ses objectifs, est « Game of Words⁶ » (Loiseau, Hallal et Ballot, 2016). Ce jeu de « devinettes » est lui tourné vers le pôle *fluency*, centré sur les activités discursives (Portine, 2013 : 162). Entre autres activités attendues du joueur, il s'agira ici d'effectuer un enregistrement audio permettant de faire deviner un mot aux autres joueurs en respectant des règles énonciatives (dans la première version, celles du jeu *Taboo*), ce qui demande d'avoir une ressource lexicale qui associe un mot à d'autres entrées utiles pour le définir (mots interdits). Dans le cadre de ce jeu, l'une des passerelles anticipées est la possibilité pour un joueur de privilégier les mots qui figurent déjà dans son dictionnaire personnel afin de lui fournir un autre point de vue sur celui-ci. Par exemple, « Anne » pourrait demander à jouer sur les mots de son lexique personnel.

Enfin, un dernier exemple de jeu et d'intégration spécifique serait celui de *Kanji Crunch*, un jeu d'association, dont il n'existe pas de prototype à l'heure actuelle contrairement aux deux autres. Celui-ci vise à permettre au joueur d'appréhender globalement l'économie du système graphique en mandarin (Goudin et Lê, 2016). Une fois les premiers niveaux joués et les compétences de base acquises (Chaix et al., 2016), un système comme LexInnova permettrait au joueur de se confronter à des niveaux construits à partir des sinogrammes issus de son lexique de groupe (cf. scénario espagnol et section 4.2.3). Mais, par le biais de cette question du sinogramme, le mandarin pose également des questions spécifiques quant à la structure du lexique.

Conclusion

À travers les exemples proposés ci-dessus, nous avons tenté de montrer d'une part la diversité des besoins en termes de structuration du lexique, mais également les ponts qui pourraient être dressés entre différents travaux menés indépendamment du fait des fonctionnalités nécessaires à la prise en compte de ces besoins. Ceux qui ont été présentés ici sont avant tout monolingues, mais d'autres (voir par exemple le projet Check Your SMILE (Yassine-Diab, Alazard-Guiu et Loiseau, 2016)), présentent également des besoins de traduction.

Au-delà des exemples ci-dessus, la plupart des équipes du projet Innovalangues développe des contenus dans plusieurs langues, ce qui nécessite tout un outillage pour la gestion du lexique. Dans la suite de l'article, nous présentons nos propositions pour un tel outil.

⁵ <http://gamer.innovalangues.net/magicword>

⁶ <http://gamer.innovalangues.net/gameofwords>

4 Conception d'une base lexicale multilingue

Bien que la gestion des développements et déploiement de l'ENPA, et donc la coordination des développements informatiques, soit centralisée, les produits issus de chaque lot sont conçus et développés indépendamment, le lien se faisant par l'intégration à l'ENPA. Mais les ressources linguistiques comme les lexiques et les corpus, d'une très fine granularité, sont structurés de façon complexe et très spécifique, et leurs tailles posent le problème du passage à l'échelle, qui doit être traité au niveau de chaque chantier.

Le projet Innovalangues est multilingue, puisqu'il permet aux élèves de s'inscrire et gérer leur formation dans plusieurs langues. Chaque langue peut sembler apprise indépendamment quand on ne considère que l'inscription d'un élève à un parcours. Mais le profil de l'élève et son histoire personnelle révèlent les spécificités (caractéristiques) multilingues de son rapport aux ressources linguistiques. De plus, les travaux en intercompréhension menés de longue date au LIDILEM (Carlo et al., 2015 ; Dabène, 1994) et intégrés aux formations du LANSAD⁷ soulignent à nouveau les besoins de ressources lexicales multilingues centralisées.

4.1 Macrostructure étendue

Les lots Innovason et GAMER exploitent tous un ou des lexique(s) réalisé(s) dans une base de données dédiée interne au lot. La diversité des besoins aurait demandé un long travail de coordination qui aurait trop ralenti chacun des chantiers, mais à l'heure où la question d'un outil centralisé pour la gestion du lexique se pose, la collecte des microstructures de chaque lexique (formes fléchies, lemmes et traits morphologiques associés⁸, transcription phonétique⁹, définitions L1 et L2¹⁰,¹¹, syllabes, accentuation¹², sinogrammes¹³, exemples en contexte¹⁴, etc.) a conduit à concevoir une structure de base lexicale multivolumes (cf. figure 5).

Les besoins d'accès par la forme fléchie conduisent à l'identification d'un volume pour celles-ci. L'intégration des transcriptions phonétiques doit être reliée aux formes fléchies. Cela conduit aussi à l'identification d'un volume dédié.

⁷ <http://lansad.u-grenoble3.fr/version-francaise/formations-en-langues/intercomprehension-en-langues-romanes-77120.kjsp>

⁸cf. Magic Word

⁹cf. InnovaSon

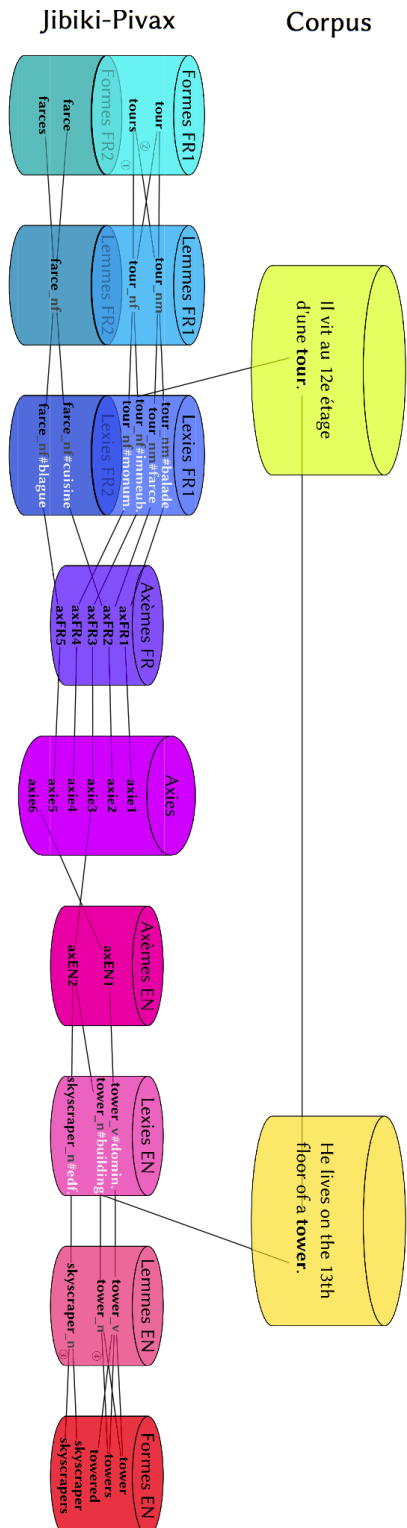
¹⁰cf. le scénario espagnol ci-dessus

¹¹cf. GAMER

¹²cf. InnovaSon

¹³cf. scénario mandarin ci-dessus

¹⁴cf. le scénario espagnol ci-dessus



Prenons l'exemple de la graphie "tour" en français. Sa transcription phonétique sera /tur/. La forme considérée ("tour") peut-être un nom masculin, "tour_nm", ou un nom féminin, "tour_nf", gérés dans le volume dédié aux lemmes. Mais ces deux lemmes peuvent avoir des sens différents, que l'on peut préciser avec une glose, comme :

1. tour_nm#potier, tour_nm#balade, tour_nm#perimètre, tour_nm#magic, tour_nm#farce,
2. tour_nf#contrôle, tour_nf#PC, tour_nf#immeuble, tour_nf#monument.

Ces différents sens sont portés par les lexies, qui sont gérées par un volume dédié. Une lexie est donc liée à une forme pour porter un sens précis.

Il est possible qu'un autre lemme conduise au même sens. Sa lexie sera différente mais le sens "unique". Les sens partagés par plusieurs lexies sont les axèmes. Un volume dédié les stocke. Disons que dans notre exemple, "tour_nm#immeuble" soit identifié par l'axème "id-axm-3856900".

Une fois le sens monolingue ainsi précisé, pour passer au même sens dans les autres langues (et pas seulement une autre langue), on définit les axes. Formellement, une axie est une classe d'équivalence de lexies synonymes (Boitet, Mangeot, Sérasset, 2002).

Si notre axème "id-axm-3856900" est relié à l'axie "id-axi-98663940", la liaison de cette axie avec les axèmes de l'anglais, gérés dans le volume des axèmes de l'anglais, pourrait conduire à identifier l'axème "id-axm-1001123", qui est lié à la lexie "tower_n#building" mais aussi à la lexie "tower_n#skyscraper". Ces lexies conduisent à deux gloses : les noms "tower_n" et "skyscraper_n". Chacun de ces noms ont deux formes, l'une au singulier, l'autre au pluriel. Les différents volumes sollicités sont présentés dans la figure 5.

Figure 1: Proto macrostructure de la base lexicale

On peut créer une telle base lexicale dans un logiciel comme Jibiki, mais :

- d'une part, il faudra étendre les fonctionnalités offertes par l' API ;
- et d'autre part, les requêtes reliant une forme dans une langue avec les formes correspondantes dans une autre langue seront coûteuses en volume. En effet, d'un point de vue conceptuel au sens des modèles entités-associations pratiqués en conception de bases de données, les lexies réalisent une association n-m entre les lemmes et les axèmes, de même pour les axes entre les axèmes des langues considérées. Il y a encore des associations n-m entre les lemmes et les formes, et entre les formes et les transcriptions phonétiques.

Les associations n-m augmentent les possibilités de parcours de façon multiplicative, même si certaines peuvent se compenser. Il en résulte un risque de temps de réponse intolérable pour les utilisateurs dans le cadre de leur tâche d'apprentissage, pour des requêtes sollicitant la base lexicale dans toute la largeur de ses volumes. Toutefois, les régularités de la langue diminueront probablement grandement la combinatoire et on pourra travailler sur des sous-volumes pour minimiser les tailles intermédiaires.

4.2 Portée des différents types de ressources

Les scénarii proposés mettent en relief des statuts différents de certaines données de la base lexicale :

- données relatives à l'utilisateur (choisies explicitement ou collectées du fait de son activité) ;
- données relatives à des groupes d'utilisateurs ;
- données issues de ressources externes.

Ceci nous a forcés à formaliser ces statuts, définissant ainsi une macrostructure fonctionnelle pour notre prototype.

4.2.1 Dictionnaire de référence

Le choix du terme **dictionnaire de référence** provient des entités (objets) qui serviront à peupler le contenu de la base lexicale. Par transitivité, on pourra parler de « dictionnaire de référence » (concept) pour parler des informations contenues dans la base lexicale et issues des dictionnaires de référence-objets. Dès lors, se pose la question du statut des informations présentes dans la base lexicale qui ne sont pas issues des dictionnaires de référence (objets). Il semble cohérent de considérer les modifications de la base lexicales validées comme intégrées au dictionnaire de référence (concept). On en arrive à la définition suivante :

dans le cadre de LexInnova, le dictionnaire de référence est la somme des informations de la base lexicale, qui a été validée, que ce soit par le biais d'une institution externe à l'instance de LexInnova (cf. dictionnaires existants) ou par le biais du système.

4.2.2 Lexique institutionnel

Un **lexique institutionnel** a pour vocation de synthétiser le point de vue sur le lexique d'une institution. Une institution pourra être tout à fait externe à LexInnova (ex : Conseil de l'Europe) ou un groupement d'utilisateurs de LexInnova. Ce lexique constitue une forme de référentiel, il n'a pas vocation à évoluer selon le déroulement d'un cours. On pourra en revanche créer un lexique institutionnel qui traduise le déroulement générique de cours dans une institution (cf. lexique cible dans le scénario « mandarin »).

4.2.3 Lexique de groupe

Un **lexique de groupe** est un lexique dynamique qui peut être modifié par un ensemble d'utilisateurs. La configuration de base sera un enseignant propriétaire d'un lexique de groupe, qui y donne accès (y compris en écriture) à un groupe d'apprenants. Toutefois, on peut tout à fait imaginer un groupe d'apprenants se créant entre eux un lexique de groupe (par exemple pour l'argot). Cela posera à terme la question de la visibilité du lexique : un lexique de groupe doit-il être visible pour les autres groupes ? Un lexique créé par un groupe d'apprenant doit-il être consultable par les enseignants ?

Le lexique de groupe permet également de formuler des objectifs pour les membres du groupe. Chaque entrée d'un lexique de groupe pourra être taguée avec la valeur cible. Ainsi le lexique cible d'un groupe pourra évoluer tout au long de l'année en fonction des ajouts de ses membres ou d'imports issus de lexiques institutionnels ou personnels.

4.2.4 Lexique personnel

Un **lexique personnel** est un lexique qui n'a qu'un seul propriétaire. Il s'agira en général d'un apprenant, qui seul pourra modifier son contenu et décider de sa visibilité. Ici aussi le propriétaire du lexique pourra choisir les éléments cibles du lexique. Au sein de ce lexique personnel, les données pourront être étiquetées pour faire la différence entre les entrées que l'apprenant a explicitement ajoutées à son lexique et celles qui y ont été ajoutées pour une raison ou pour une autre par le système (cf. scénarios « jeu »). Quoi qu'il en soit, l'apprenant membre d'un groupe devra décider lui-même si ses objectifs coïncident avec ceux du groupe (des groupes) au(x)quel(s) il appartient. En d'autres termes, ce n'est pas parce qu'une entrée figure dans un lexique de groupe auquel a accès un apprenant qu'elle se trouvera dans son lexique personnel.

5 Résultat fonctionnel

Si les étapes précédentes nous ont permis d'affiner la modélisation du lexique, des questions restent en suspens (cf. partie précédente). Afin d'alimenter les discussions avec les futurs utilisateurs et d'orienter nos recherches, nous avons décidé d'implémenter rapidement un prototype fonctionnel. Celui-ci sert également à décider des efforts de développement futurs. Ce prototype est développé en PHP de façon à pouvoir être intégré facilement à l'ENPA développé dans le projet Innovalangues. Il se sert de la plate-forme Jibiki pour la gestion des ressources lexicales. La communication entre les deux systèmes s'effectue via l'interface de programmation de Jibiki fondée sur le protocole REST.

Jibiki (Mangeot & Chalvin, 2006) est une plate-forme générique de gestion de ressources lexicales hétérogènes en ligne développée et maintenue depuis 2001 principalement par Mathieu Mangeot. Celle-ci est programmée à l'aide de Enhydra, un serveur d'objets java basé sur une architecture 3/tiers. Pour la couche de données, la base de données utilisée est Postgres. Le logiciel est disponible gratuitement et en source ouverte sur la forge du LIG¹⁵.

Tout type de ressource lexicale au format XML peut être importé dans la plate-forme. Les ressources sont décrites et manipulées à l'aide de pointeurs dans la structure XML, ce qui permet de les importer, les consulter et les éditer sans modifier leur structure. Il est possible de gérer tout type de microstructure. Concernant la microstructure, le système de traitement de liens inter-volumes permet de gérer différentes macrostructures : monolingue, bilingue bidirectionnelle, multilingue à structure pivot ou multi-étages type Pivax (Zhang et al., 2014).

5.1.1 Consultation des ressources

Deux interfaces de consultation sont disponibles. L'interface de consultation simple affiche une vue évoluée du dictionnaire imprimé : la partie gauche affiche les vedettes du voisinage immédiat du

¹⁵<https://jibiki.ligforge.imag.fr/>

mot recherché classées par ordre alphabétique. Elle est équipée d'un ascenseur infini permettant de parcourir toutes les vedettes du dictionnaire selon l'ordre alphabétique. Lorsqu'on clique sur une vedette de la partie gauche, l'article complet s'affiche dans la partie droite.

L'interface de consultation avancée permet de combiner les critères de recherche. Par exemple, rechercher un mot français terminant par "er", qui soit un nom et qui appartienne au domaine de la botanique.

Lors de l'affichage de chaque article, un menu s'affiche en haut à droite. Celui-ci comprend des liens vers le formulaire d'édition, l'historique des modifications, ainsi que la vue source XML de l'article.

5.1.2 *Édition des articles*

Lors de l'import d'une ressource sur la plate-forme, le schéma XML représentant la structure des articles est également importé. Un formulaire d'édition des articles est ensuite automatiquement généré par la plate-forme (Mangeot & Chalvin, 2006).

Deux modes d'édition sont possibles. Lors de la consultation d'un article, l'édition sur place permet de modifier directement le texte d'un article qui s'affiche en double-cliquant sur la chaîne de caractères à modifier. Celle-ci se transforme en champ de texte avec un bouton « ok » sur la droite pour valider la saisie.

Cet éditeur est programmé à l'aide de la technologie AJAX et il utilise l'API REST de la plate-forme Jibiki pour dialoguer avec le serveur. Lors de la validation de la saisie (clic sur le bouton « ok »), la nouvelle chaîne de caractères est envoyée au serveur avec le pointeur Xpath du segment édité et l'identifiant unique de l'article (Mangeot, 2016).

Par contre, ce mode ne permet pas l'ajout de nouvelles informations, comme un exemple d'usage en contexte. Pour ajouter ou supprimer des parties d'information, il est nécessaire d'utiliser le formulaire d'édition complète. Outre les interacteurs HTML classiques (boîtes texte, menu déroulant, cases à cocher, etc.), celui-ci a été enrichi avec des interacteurs plus complexes pour gérer des listes d'objets.

5.2 **Présentation du système Lexinnova**

5.2.1 *Architecture globale du système*

Le site Web principal avec lequel l'utilisateur interagit est l'ENPA du projet Innovalangues.

L'ENPA est déjà intégré à la plate-forme Claroline Connect. Le prototype Lexinnova¹⁶ reprend la charte graphique de l'ENPA et il est en cours d'intégration sous forme de module Claroline Connect.

Les fonctionnalités du prototype sont directement intégrées à l'ENPA sous forme de menus en haut à droite de l'écran.

Une instance spécifique de Jibiki a été déployée pour le projet Lexinnova.

Celle-ci contient :

- un dictionnaire de référence Lexinnova avec un volume pour chaque langue (pour l'instant seul l'espagnol est accessible) ;
- un dictionnaire créé pour chaque utilisateur avec un volume pour chaque langue.

Cette instance est installée sur le serveur de démonstration de l'équipe GETALP, au Laboratoire d'Informatique de Grenoble. La communication entre le prototype et l'instance de Jibiki s'effectue avec l'interface de programmation (API) REST de la plate-forme Jibiki. La figure 2 montre l'architecture du système dans son ensemble.

¹⁶<http://totoro.imag.fr/Lexinnova>

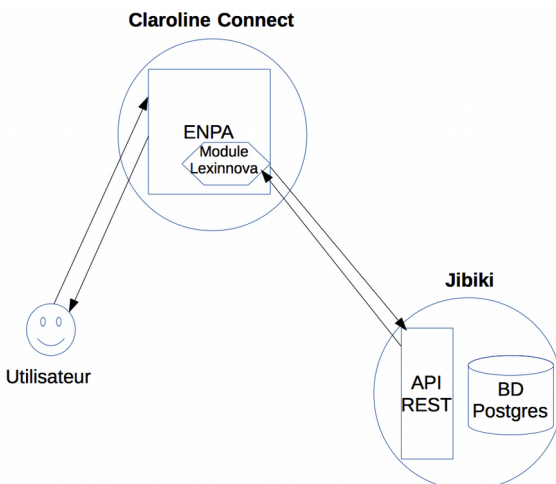


FIGURE 2: Architecture du système Lexinnova

5.3 Fonctionnement du prototype

Lors de la première visite d'un utilisateur sur l'ENPA et avant que son compte ne soit créé, deux fonctionnalités sont disponibles dans le menu Lexinnova :

- consultation des dictionnaires de référence ;
- lecture active.

Lorsque l'utilisateur crée un compte sur l'ENPA, un compte similaire est créé sur la plate-forme jibiki via un appel à l'API.

Une fois le compte créé, une nouvelle fonctionnalité est ajoutée au menu : la gestion des lexiques personnels (consultation et modification).

Si l'utilisateur a des droits spécifiques d'administration (par exemple, un enseignant ou un administrateur de l'ENPA), un second menu d'administration est ajouté à la barre des menus.

Il contient les fonctionnalités suivantes :

- gestion des utilisateurs (liste de tous les utilisateurs et suppression d'un utilisateur existant);
- gestion des lexiques (liste de tous les lexiques, création d'un nouveau lexique et suppression d'un lexique existant).

Une démonstration de ce prototype a été réalisée lors du séminaire bi-annuel Innovalangues de janvier 2016¹⁷. Un podcast est à présent disponible¹⁸.

5.4 Fonctionnalités disponibles

5.4.1 Limitations du prototype

De façon à pouvoir rapidement montrer un prototype fonctionnel, nous avons choisi de limiter les fonctionnalités disponibles en nous concentrant sur certains scénarios (scénario « tfo »). La réflexion

¹⁷ <http://innovalangues.fr/seminaire-bi-annuel-innovalangues-janvier-2016/>

¹⁸ <http://podcast.grenet.fr/episode/lexinnova-lexiques-personnalisables-integres-a-lenpa/>

sur les lexiques de groupe étant encore en cours, nous n'avons pas implémenté cette fonctionnalité. Les dictionnaires de références et les lexiques personnels ont été quant à eux complètement implémentés.

Au niveau des données, nous n'avons pas encore collecté un ensemble complet de données lexicales pour chaque langue du projet. Nous avons créé un dictionnaire de référence pour l'espagnol afin d'implémenter certains scénarios. Ce dictionnaire ne contient pour l'instant que le vocabulaire nécessaire à la réalisation des scénarios (soit une vingtaine de mots au total).

Pour le passage à l'échelle en termes de quantité de données et de nombre d'utilisateurs, des calculs ont été faits sur la base de 180 000 entrées pour chaque dictionnaire de référence et 10 000 utilisateurs inscrits. Des tests de montée en charge ont été effectués sur la base de 600 utilisateurs en ligne. Le prototype actuellement déployé peut répondre à une telle charge pour les fonctionnalités principales de gestion des ressources. Par contre, le module de lecture active nécessitant beaucoup de mémoire vive, il sera nécessaire de le déployer sur un autre serveur afin de répartir la charge.

5.4.2 Consultation des dictionnaires de référence

La consultation des dictionnaires de référence est publique. Il n'est donc pas nécessaire de se connecter pour y accéder. L'interface utilisée est la page de consultation simple de la plate-forme Jibiki : la partie gauche affiche les vedettes du voisinage immédiat du mot recherché et la partie droite affiche le ou les articles trouvés de manière détaillée. Si l'utilisateur est connecté, un bouton "+" s'affiche en haut à droite de l'article. L'utilisateur peut cliquer sur ce bouton pour importer l'article dans son lexique personnel.

La figure 3 montre la fenêtre de consultation du mot "tio" dans le dictionnaire espagnol (on notera

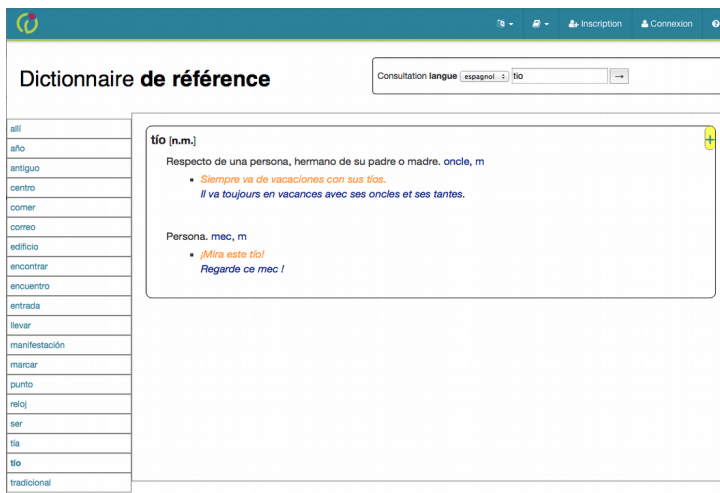


FIGURE 3: Consultation du dictionnaire de référence

que l'article « tío », accentué, est obtenu).

5.4.3 Gestion des lexiques personnels

FIGURE 4: Édition d'un exemple d'article dans le lexique personnel5



La gestion des lexiques personnels reprend également l'interface de consultation simple de la plateforme Jibiki. Il est par contre nécessaire d'être un utilisateur enregistré pour pouvoir y accéder. Lorsque l'utilisateur consulte une entrée dans cette interface, il peut ensuite l'éditer directement en cliquant sur la zone à modifier. La technique utilisée est celle de l'édition sur place de la plate-forme Jibiki. Sur la figure 4, l'utilisateur a consulté le mot "tío" dans son lexique personnel (après l'avoir importé depuis le dictionnaire de référence). Il souhaite modifier l'exemple pour qu'il fasse référence à son histoire personnelle ce qui lui permettra d'accéder plus facilement au sens de ce mot.

5.4.4 Module de lecture active

Le module de lecture active reprend celui utilisé sur le site jibiki.fr (Mangeot 2016). Il s'agit d'une aide à la lecture pour les apprenants d'une langue étrangère. L'utilisateur saisit ou colle un texte dans la fenêtre de saisie. Le texte est ensuite envoyé à un analyseur morphologique puis chaque lemme est consulté dans son lexique personnel ou dans le dictionnaire de référence si aucune réponse n'est trouvée dans son lexique personnel. Le texte entré précédemment est réaffiché plus bas avec des informations supplémentaires provenant de l'analyse. Lorsque l'utilisateur bute sur un mot, il clique dessus et une fenêtre s'affiche avec l'article correspondant provenant de la consultation des différentes ressources (lexique personnel ou dictionnaire de référence).

Dans l'exemple de la figure 5, l'utilisateur clique sur le mot "antiguo" et l'article s'affiche sur la droite de l'écran. Si l'article provient du dictionnaire de référence, celui-ci affichera un bouton "+" permettant d'importer l'article dans le lexique personnel de l'utilisateur, comme lors de la consultation du dictionnaire de référence.

Le module de lecture active peut également afficher une transcription ou une prononciation au dessus de chaque mot en utilisant l'élément ruby¹⁹ du standard HTML5. Pour le japonais, il est possible d'afficher le furigana ou le romaji. Pour le chinois, il est possible d'afficher le pinyin ou le bopomofo. Pour le français, il est possible d'afficher une prononciation simplifiée ou en API, etc.

¹⁹ <http://www.w3.org/TR/ruby/>

Lecture active : saisissez un texte ?

Es el centro neurálgico de Madrid. Es un punto de encuentro, de manifestaciones... Allí se encuentra el reloj del antiguo edificio de Correos que marca tradicionalmente la entrada del año.

español - Analyser le texte

Note : cliquez sur un mot pour afficher ses informations.

Es el centro neurálgico de Madrid. Es un punto de encuentro, de manifestaciones. antiguo edificio de Correos que marca tradicionalmente la entrada del año.

antiguo [adj.]

Que existe desde hace mucho tiempo.
vieux, m; antique, m.
Esa taberna es muy antigua. Cette taverne est très vieille.

FIGURE 6: Module de lecture active

6 Conclusion

La conception d'un ENPA a fait émerger la nécessité d'une grande base lexicale commune couvrant tous les besoins des différents acteurs, que se soient des machines (modules de jeux sérieux ou de générateurs d'exercices, etc.) ou des humains (apprenants). À partir de plusieurs scénarios d'utilisation, nous avons défini une structure de base lexicale multilingue permettant de répondre à ces besoins. Nous avons ensuite implémenté un prototype fonctionnel permettant la consultation automatique ou manuelle de ressources et la création de lexiques personnels. Ce prototype, programmé en PHP, interagit avec la plate-forme Jibiki de gestion de ressources lexicales.

Les perspectives principales de ce projet à moyen terme sont les suivantes :

- l'analyse des pratiques d'enseignants en langues en termes de lexiques grâce à un questionnaire en cours d'élaboration ;
- le passage à l'échelle du prototype en termes de nombre d'utilisateurs, de quantité de données et de macrostructure. Pour cela, il faudra constituer pour chaque langue un volume de grande couverture avec des données libres de droit (possibilité d'utiliser les données de wiktionary) et de relier les différents sens de chaque langue via un volume pivot ;
- un test grandeur nature au deuxième semestre 2016-2017 avec une classe d'étudiants en sciences apprenant l'anglais de spécialité à l'Université de Grenoble.

Remerciements

Cette recherche a été en grande partie financée par le projet IDEFI Innovalangues.

Références

BASTIEN, C. ET SCAPIN, D. (2004). « La conception de logiciels interactifs centrée sur l'utilisateur: étapes et méthodes », in Pierre Falzon (dir.), *Ergonomie*, 1^{re} édition, Paris, Presses Universitaires de France, pp. 451–462.

- BELLASSEN, J. (2010). « La didactique du chinois et la malédiction de Babel : émergence, dynamique et structuration d'une discipline » in *Études Chinoises*, Hors-série. Disponible en ligne : http://www.afec-etudeschinoises.com/IMG/pdf/Bellessen_Didactique.pdf.
- BOITET, C., MANGEOT, M., SÉRASSET, G. (2002), « The PAPILLON project: cooperatively building a multilingual lexical data-base to derive open source dictionaries & lexicon », Proc of on NLP and XML (NLPXML 2002), Ed. Graham Wilcock, Nancy Ide & Laurent Romary, COLING Workshop, Taipei, Taiwan, 31 August 2002, pp. 9-15.
- CARLO, M.D., ANQUETIL, M., VECCHI, S., JAMET, M.-C., MARTIN, E., PEREA, E.C., HIDALGO, R., PISHVA, Y., GILLES, F. ET ANDRADE, A.I. (2015). « REFIC – Référentiel de compétences de communication plurilingue en intercompréhension », Référentiel de compétences Projet Européen Miriadi. Disponible en ligne : <http://www.miriadi.net/refic>.
- CARRON T., MARTY J-C. ET MANGEOT M. (2009). *How to bring immersion in Learning Games ?* Proc. IEEE-ICALT 2009, Riga, Latvia, 15-17 July 2009, 6 p.
- CECRL (2000). *Cadre européen commun de référence pour les langues : apprendre, enseigner, évaluer*, Conseil de la coopération culturelle — Comité de l'éducation — Division des langues vivantes (dir.), Édition française, Strasbourg; Paris, Conseil de l'Europe. Disponible en ligne : <http://medias.didierfle.com/media/contenuNumerique/007/4140016745.pdf>.
- CHAIX, C., MEUNIER-CARUS, M., PESQUET, J.-A., RONDIN, R., GOUDIN, Y. ET LOISEAU, M. (2016). « Kanji Crunch: Apprentissage dissocié des habiletés et autonomisation de la compétence graphique par le jeu vidéo en mandarin langues étrangères », 38e colloque APLIUT : *Jeux en jeu dans l'enseignement/l'apprentissage des langues en LANSAD*, Lyon, 2-4 juin 2016.
- DABÈNE, L. (1994). « Le projet européen GALATEA: pour une didactique de l'intercompréhension en langues romanes » Jeanine Stolidi (dir.), *Recherches en linguistique hispanique*, n°22, pp. 41-45.
- GOUDIN, Y. ET LÊ, T.M. (2016). « Jouer avec le sacré? Le sinogramme à l'ère du jeu sérieux » Haydée Silva et Mathieu Loiseau (dir.), *Recherches et applications*, n°59, pp. 145-160.
- LOISEAU, M., HALLAL, R. ET BALLOT, P. (2016). « Game of Words: prototype of digital game focusing on oral production (and comprehension) through asynchronous interaction », EUROCALL 2016, Limassol (Chypre), 24-27 août 2016.
- LOISEAU, M., ZAMPA, V. ET REBOURGEON, P. (2015). « Magic Word: premier jeu développé dans le cadre du projet Innovalangues », *ALSIC*, vol. 18, n°2. DOI : 10.4000/alsic.2828. Disponible en ligne : <http://alsic.revues.org/2828>.
- LUSTE-CHAA, O. (2009). *Les acquisitions lexicales en français langue seconde: conceptions et applications*, Thèse de doctorat, Université Paul Verlaine, Metz. Disponible en ligne : <http://www.theses.fr/2009METZ023L>.
- MANGEOT M. (2016). *Collaborative construction of a good quality broad coverage and copyright free Japanese-French dictionary* Hosei University International Found Foreign Scholar Fellowship Report Volume XVI 2013-2014, Hosei University, Tokyo, Japan, pp. 175-208.

MANGEOT M. ET CHALVIN A. (2006). *Dictionary Building with the Jibiki Platform : the GDEF case*. Proc. of LREC 2006, Genoa, Italy, 23-25 May 2006, pp. 1666-1669.

MASPERI, M. ET QUINTIN, J.-J. (2014a). « Enseigner à l'université en France, à l'ère du numérique: l'apport de dispositifs d'ingénierie innovants dans la formation en langues », in Cristiana Cervini et Anabel Valdivieso (dir.), *Dispositivi formativi e modalità ibride per l'apprendimento linguistico*, Bologna, CLUEB, Contesti Linguistici, pp. 61-80. Disponible en ligne : <https://www.researchgate.net/publication/271852910>.

MASPERI, M. ET QUINTIN, J.-J. (2014b). « L'innovation selon Innovalangues » Elsa Del Col (dir.), *Lingua e nuova didattica*, n°1/2014, pp. 6-14.

PORTINE, H. (2013). « L'ingénierie linguistique : des technologies au service d'une didactique intégrant la cognition ? » Christian Ollivier et Laurent Puren (dir.), *Mutations technologiques, nouvelles pratiques sociales et didactique des langues*, n°54, pp. 159–168.

YASSINE-DIAB, N., ALAZARD-GUIU, C. ET LOISEAU, M. (2016). « Check your Smile, first prototype of a collaborative LSP website for technical vocabulary learning », EUROCALL 2016, Limassol (Chypre), 24-27 août 2016.

ZHANG Y., MANGEOT M., BELLYNCK V. ET BOITET C. (2014). *Jibiki-LINKS: a tool between traditional dictionaries and lexical networks for modelling lexical resources*. Proceedings of the 4th Workshop on Cognitive Aspects of the Lexicon (CogALex) 2014 (Eds. Michael Zock, Reinhard Rapp, Chu-Ren Huang), Dublin, Ireland, 23 August 2014, 12 p.