Contributions et corrections dans le dictionnaire japonais- français jibiki.fr

Mathieu Mangeot, Mutsuko Tomokiyo

Contributions et corrections dans le dictionnaire japonais-français jibiki.fr

Mathieu Mangeot & Mutsuko Tomokiyo
Laboratoire LIG, équipe GETALP
Bâtiment IMAG CS 40700
38058 GRENOBLE CEDEX 9, FRANCE
{mathieu.mangeot,mutsuko.tomokiyo}@imag.fr

Résumé

Concernant le couple de langues français-japonais, les ressources disponibles sur le Web sont peu nombreuses et de taille modeste. Il existe cependant de nombreux dictionnaire imprimés de qualité et à large couverture. C'est pourquoi nous avons lancé le projet jibiki.fr de construction d'un dictionnaire japonais-français de qualité et à large couverture à partir de récupération de données issues du dictionnaire de Jean-Baptiste Cesselin (1940) que nous avons numérisé et lu optiquement. Nous avons complété ces données par d'autres issues du dictionnaire JMdict de Jim Breen (2004) et de Wikipedia pour pallier le manque de vocabulaire récent. Nous avons ensuite installé ces données sur une plateforme de gestion de ressources lexicales en ligne. Les utilisateurs peuvent alors consulter le dictionnaire et corriger les erreurs qu'ils trouvent lors de la consultation. La plateforme possède également une interface de programmation (API) qui permet de programmer des scripts afin de corriger automatiquement certains phénomènes repérés lors de consultations. La ressource ainsi construite est disponible en téléchargement libre de droits. Cet article décrit certaines techniques de correction automatique et manuelle à l'issue des 3 ans et demi du projet.

Contributions in the jibiki.fr Japanese-French dictionary

Concerning the French-Japanese language couple, the resources available on the Web are scarce and modest in size. However, there are many high-quality printed dictionaries. This is why we launched the jibiki.fr project to build a high-quality, wide-coverage Japanese-French dictionary based on the recovery of data from Jean-Baptiste Cesselin's (1940) dictionary, which we digitized and ocerized. We added other data from Jim Breen's (2004) JMdict dictionary and Wikipedia to compensate the lack of recent vocabulary. We then installed this data on an online lexical resource management platform. Users can then consult the dictionary and correct any errors they find during the consultation. The platform also has a programming interface (API) that allows you to program scripts to automatically correct certain phenomena identified during consultations. The resulting resource is downloadable online copyright free. This article describes some automatic and manual correction techniques at the end of the 3.5 years of the project.

日仏辞書 jibiki.fr における利用者協力と修正

日仏言語対はウエブ上で使える資源が少なく、その規模も小さい。しかしながら書籍版の良い多岐な情報を含む辞書が存在する。それ故に我々は、ジャン・バプチスト・セスラン(1940) の辞書から　高質な日仏辞書の作成を目指す jibiki.fr の企画を立てた。 セスラン辞書をコード化し光学的に読み取った。さらに新しい語彙を補うために、ジム・ブリーン (2004)の JMdict 辞書と Wikipedia のデータ　を加え、コンピュータで管理できるように開発したプラットホームに置いた。プラットホームは利用者が辞書を引くことができると同時に、辞書引きの際に見いだした誤りを人手または自動的に修正することができるプログラム(API) を搭載している。こうして作られた資源は自由にダウンロードすることができる。本稿は３年半にわたる自動あるいは人手による修正の技術を報告する。

Mots-clés : dictionnaire japonais-français, correction d'erreurs, OCR, numérisation de dictionnaires, annotations et classification d'exemples

Keywords : Japanese-French dictionary, error detection and correction, OCR, examples classification and annotation

1. Introduction

Bien que le français et le japonais soient considérés comme des langues bien dotées concernant les outils et les ressources linguistiques, le couple franco-japonais est considéré comme une paire de langues peu dotée en ce qui concerne sa disponibilité sur le Web. En effet, il existe peu de ressources lexicales électroniques bilingues de qualité, et à la fois gratuites et libres de droit. Les corpus bilingues franco-japonais et systèmes de traduction automatique sont également rares.

Heureusement, il existe des dictionnaires imprimés français-japonais de bonne qualité et suffisamment anciens pour être libres de droits, tels que le dictionnaire japonais-français de Gustave Cesselin (Cesselin, 1940). Nous avons réutilisé cette ressource pour construire un dictionnaire de bonne qualité et de large couverture, et le rendre disponible sur le Web (Mangeot, 2016). Le dictionnaire Cesselin a d'abord été scanné, puis lu optiquement et analysé pour détecter les mots-clés et les articles. Ensuite, plusieurs corrections d'erreurs ont été effectuées sur le français et le japonais. Afin de mettre à jour ces données dont le vocabulaire est parfois ancien, nous avons réutilisé des ressources électroniques existantes telles que Wikipedia et le dictionnaire électronique japonais-anglais Jmdict (Breen, 2002). La ressource résultante a été ensuite mise en ligne sur le Web [1] pour consultation et correction par des contributeurs volontaires. Les données sont disponibles dans le domaine public.

Cette méthodologie pourrait être appliquée à d'autres couples de langues dans une situation similaire avec de bons dictionnaires imprimés mais peu de ressources électroniques.

Dans cet article, après trois ans et demi de vie du projet, nous effectuons un bilan des contributions manuelles et automatiques.

2. Description de la ressource construite

Le dictionnaire jibiki.fr est un dictionnaire bilingue monodirectionnel japonais→français au format XML. Les données sont versées dans le domaine public (licence Creative Commons CC0) et disponibles au téléchargement sur le site du projet[2]. Il contient actuellement 154 209 articles.

La microstructure des articles est dérivée de celle du Cesselin (voir Figure 1). Le mot-vedette est noté dans 4 scripts différents : romaji hepburn[3], hiragana[4] et japonais (kana+kanji)[5]. Le corps de l'article est divisé en blocs grammaticaux identifiés chacun par une classe grammaticale différente. Chaque bloc grammatical est ensuite divisé en blocs sémantiques. Chaque bloc sémantique représente un sens de mot avec ses traductions en français. La suite de l'article est composée d'une liste d'exemples, proverbes, collocations ou locutions triées selon le romaji.

2.1. Articles provenant du Cesselin

Les articles du dictionnaire Cesselin ont été repris dans leur intégralité. Il s'agit d'un total de 82 710 articles (soit 54 % du total). Les données provenant d'un processus de lecture optique, les articles contiennent un nombre important d'erreurs. Toutefois, excepté pour le mot-vedette, ces erreurs n'empêchent pas la compréhension du texte par un humain.

Le défaut majeur dans la microstructure du Cesselin réside dans les exemples. d'une part, ils sont listés à la fin de l'article sans être rattachés à un sens particulier et d'autre part, il n'est pas possible de distinguer son type (exemple d'usage, collocation, locution figée, proverbe, etc.).

1 http://jibiki.fr/
2 https://jibiki.fr/data/
3 Le romaji hepburn est une méthode de romanisation du japonais, inventée par James Curtis Hepburn en 1887.
4 Les hiragana sont un syllabaire japonais
5 Il n'est pas rare que le japonais soit écrit avec quatre systèmes d'écriture différents dans une phrase : le syllabaire katakana pour des mots d'origine étrangère, les kanji (caractère chinois), le syllabaire hiragana pour de la conjugaison de verbes et d'adjectifs et la déclinaison de noms, et le romaji pour les mots d'origine étrangère.

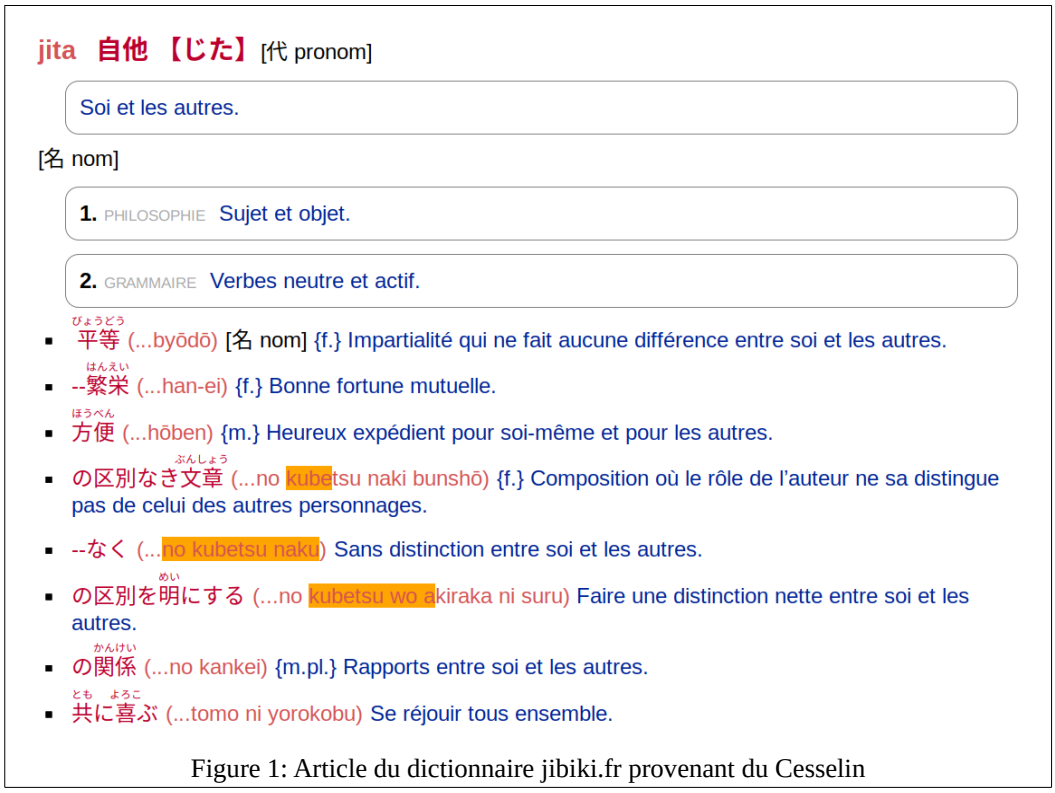

Figure 1: Article du dictionnaire jibiki.fr provenant du Cesselin

2.2. Articles provenant du JMdict

Le JMdict (Breen, 2004) est un dictionnaire bilingue monodirectionnel japonais→anglais contenant 182 128 articles pour la version du 18 février 2019. Il contient également un certain nombre d'articles traduits dans d'autres langues : allemand, espagnol, français, hongrois, russe, suédois, etc. Il est disponible gratuitement au téléchargement[6].

Nous avons repris les articles qui n'existaient pas dans le Cesselin mais qui se trouvaient dans le dictionnaire monolingue japonais Super Daijirin[7] (inclus dans MacOs). Nous avons repris au total 48 188 articles (soit 31 % du total) dont 2 667 en français et 45 521 en anglais.

La microstructure du JMdict est assez limitée. Il n'y a pratiquement pas d'exemples et les sens de mot ne sont pas partagés entre les traductions des différentes langues.

Par la suite, nous avons constaté qu'une quantité non négligeable d'articles du dictionnaire JMdict étaient en fait inclus dans le Cesselin comme des collocation d'autres articles. Nous avons le projet de les remplacer automatiquement par des renvois vers les articles existants du Cesselin.

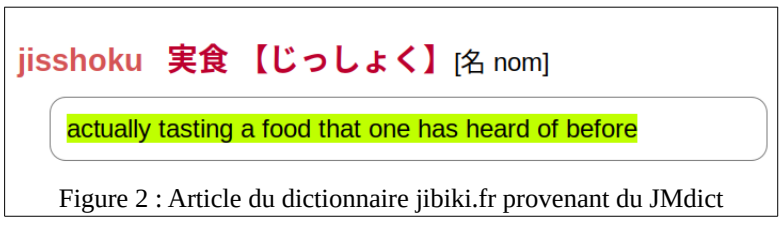

Figure 2 : Article du dictionnaire jibiki.fr provenant du JMdict

6 http://www.edrdg.org/jmdict/edict_doc.html
7 Super Daijirin 3.0 スーパー大辞林 3.0 （三省堂, Sanseidô Shoten, Tokyo, 2006）

2.3. Articles provenant de Wikipedia

Pour compléter cet ensemble de données, nous avons repris des articles de Wikipedia qui ne se trouvaient ni dans le Cesselin, ni dans le JMdict mais qui étaient attestés dans Super Daijirin[7] . Au total, nous avons importé 23 507 articles dont 20 857 en français et 2 650 en anglais.

Les articles issus de Wikipedia contiennent le mot-vedette et sa traduction en français ou en anglais ainsi qu'un lien vers les pages Wikipedia. Ils ne contiennent pas d'exemple ou de sens de mot.

Nous avons le projet d'automatiser cet import de façon à l'effectuer périodiquement. En effet, Wikipedia étant en constante évolution, les liens interlingues entre les articles de différentes langues s'enrichissent constamment.

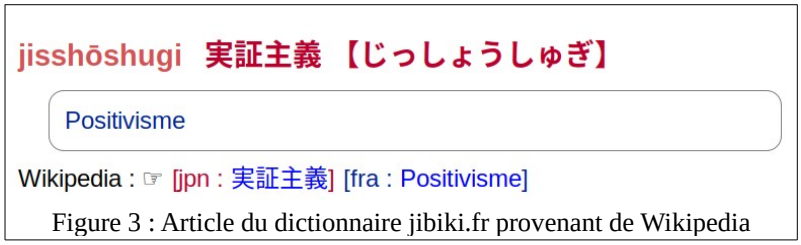

Figure 3 : Article du dictionnaire jibiki.fr provenant de Wikipedia

3. Processus de correction

Les corrections sont trop nombreuses et les contributeurs trop peu nombreux pou envisager un travail de correction systématique de toutes les données. La stratégie choisie est la suivante : mettre en place des outils permettant aux utilisateurs de contribuer très facilement lorsqu'ils consultent un mot et constatent une erreur.

3.1. Édition rapide

Certaines catégories d'information peuvent être corrigées par les contributeurs directement lors de la lecture d'un article. Il suffit de cliquer sur le segment à corriger. Celui-ci apparaîtra alors dans un champ texte et le contributeur validera ensuite sa correction en cliquant sur un bouton OK.

Les informations qu'il est possible de corriger de cette manière sont : les mots-vedette s'ils n'ont pas été validés, les traductions françaises, et les exemples (kanji, romaji et traduction française).

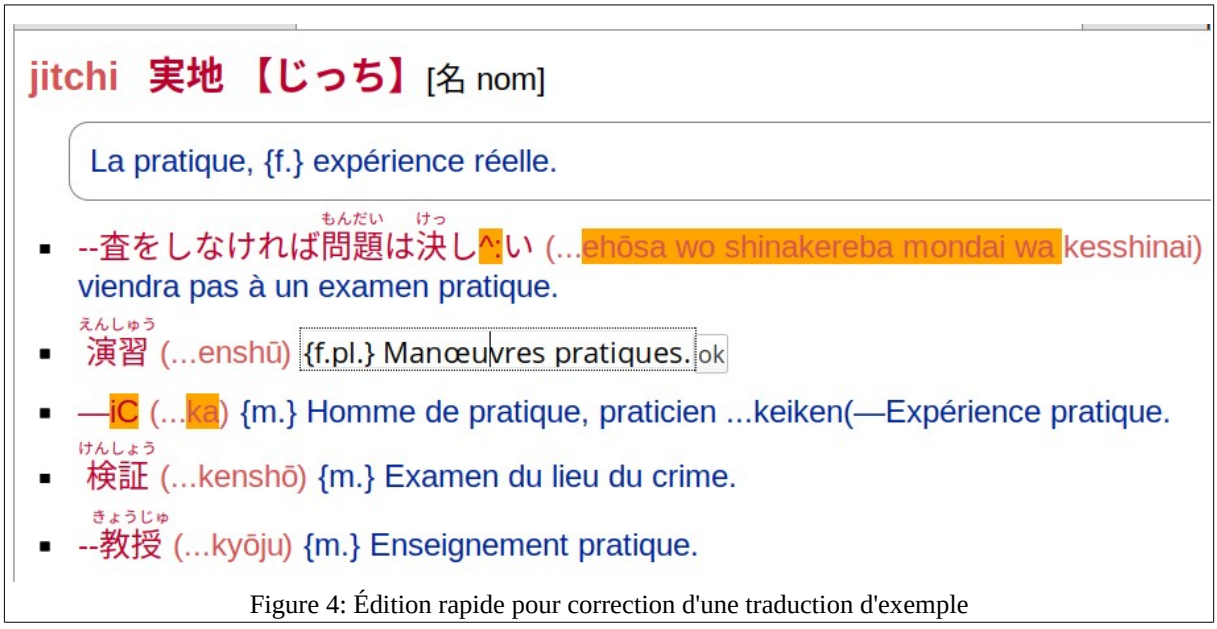

Figure 4: Édition rapide pour correction d'une traduction d'exemple

3.2. Édition complète

Il n'est par contre pas possible de modifier la structure d'un article avec l'édition rapide. Pour des manipulations plus lourdes (création d'un nouvel article, ajout d'un bloc grammatical, d'un sens ou d'un exemple), il faut utiliser le formulaire d'édition complète.

Figure 5 : Formulaire d'édition complète

Dans celui-ci, toutes les informations présentes dans l'article sont affichées dans des champs de texte. Il existe des interacteurs spécifiques (voir les boutons + et – sur la figure 3) pour ajouter ou supprimer un bloc (sens, exemple, etc.). Une fois les corrections effectuées, le contributeur accède à la vue complète de l'article modifié, tel qu'il sera affiché dans le dictionnaire, puis le valide ou non (voir Figure 3).

3.3. Tableau résumé des contributions

Au total, 78 361 modifications d'articles ont été effectuées dont :

- 10 329 mots-vedette en kanji ajoutés,
- 5 167 mots-vedette en kanji vérifiés,
- 4 390 mots-vedette en kanji corrigés,
- 4 649 traductions en français

Les 5 contributeurs les plus actifs sont :

1. Mutsuko Tomokiyo : 35 319 contributions terminées
2. Robot Dictionnaire : 21 082 contributions terminées
3. Mathieu Mangeot : 13 852 contributions terminées
4. Nicolas Mollard : 5 955 contributions terminées
5. Louis Lecailliez : 608 contributions terminées

Comme son nom l'indique, Robot Dictionnaire est un robot effectuant des corrections automatiques sur le dictionnaire.

4. Corrections automatiques

Les corrections automatiques s'effectuent à l'aide de scripts programmés en Perl qui utilisent l'interface de programmation (API) de la plateforme jibiki (voir http://jibiki.fr/jibiki/Api.po). Les étapes principales sont les suivantes :

- Une liste des identifiants de chaque article à modifier est extraite de la base et envoyée en entrée du script ;
- Pour chaque article, le script demande au serveur via l'API le code XML de l'article ;
  - ex : demande du code XML de l'article numéro 23158254
    `curl "http://jibiki.fr/jibiki/api/Cesselin/jpn/handle/23158254"`;
- Le script récupère l'identifiant de la contribution dans le code XML de l'article ;
- Le script renvoie au serveur via l'API le pointeur XPath de la partie d'article à modifier ainsi que la chaîne de caractères modifiée.
  - ex : modification du hiragana de l'article 大きい par おおきい
    `curl -X PUT -u user:password -d "/volume/d:contribution/d:data/article/forme/vedette/vedette-hiragana/text()" "http://jibiki.fr/jibiki/api/Cesselin/jpn/jpn.大きい.22491801.c/おおきい"`;

4.1. Transcription du kanji 大

En japonais, le o long est noté ō ou ô en romaji et おう [ou] en hiragana sauf lorsqu'il transcrit le kanji 大. Dans ce cas, il est transcrit おお [oo]. Par exemple, le nom 王族 (famille royale) sera transcrit ōzoku en romaji et おうぞく [ouzoku] en hiragana ; l'adjectif 大きい (grand) sera transcrit ōkii en romaji et おおきい [ookii] en hiragana.

Avant la mise en ligne du dictionnaire, lorsque pour chaque mot-vedette, nous avons généré le hiragana à partir du romaji présent dans le dictionnaire, nous n'avions pas pensé à intégrer cette exception. Il a donc fallu la corriger à l'aide d'un script automatique une fois le dictionnaire mis en ligne. 571 mots-vedettes ont été corrigés.

4.2. Okurigana

Les okurigana (送り仮名) sont des suffixes qui suivent la racine en kanji (漢字, caractères chinois) des verbes et des adjectifs. Les conventions d'écriture en okurigana ont changé à plusieurs reprises, et la dernière version a été standardisée par le gouvernement japonais en 1973. Comme le Cesselin (publié en 1939) a utilisé une ancienne convention pour les okurigana, nous les avons remplacés de manière automatique par la convention en cours en utilisant le dictionnaire Super Daijirin inclus dans MacOs. Par exemple :

- 考へる → 考える (kangaeru, penser),
- 空騒 (karasawagi, du bruit pour rien) → 空騒ぎ

Pour chaque mot-vedette du dictionnaire Cesselin, le dictionnaire Super Daijirin[8] est consulté avec le furigana (prononciation) du mot-vedette (からさわぎ dans le deuxième exemple). Ensuite, les hiragana sont supprimés du mot-vedette du Cesselin ainsi que des mots-vedette trouvés dans le Super Daijirin (空騒ぎ → 空騒 dans le deuxième exemple). Si des mots-vedette correspondent (空騒 = 空騒), alors le mot-vedette du Cesselin est remplacé (空騒 → 空騒ぎ) et l'ancienne version est gardée dans l'article avec une mention (ancien okurigana). Cela permet de retrouver l'article 空騒ぎ même si la recherche est effectuée avec l'ancienne version (空騒).

Nous avons corrigé de cette manière environ 6 500 articles.

8 Super Daijirin 3.0 スーパー大辞林 3.0 （三省堂, Sanseidô Shoten, Tokyo, 2006）

4.3. Mots-vedette

Les données provenant de lecture optique, il est nécessaire de vérifier chaque mot-vedette pour s'assurer qu'il n'y a pas eu d'erreur de reconnaissance de caractères.

Plusieurs cas ont été distingués :

- Le mot-vedette et le furigana apparaissent dans un autre dictionnaire : celui-ci est marqué vérifié. Cela concerne 42 219 mots-vedette, soit 50 % du total.
- Le mot-vedette n'apparaît pas dans d'autres dictionnaires mais il se prononce comme indiqué en furigana : celui-ci est marqué à vérifier. Cela concerne 12 922 mots-vedettes soit 16 % du total. Il reste encore 7 765 mots-vedette à vérifier.
- Le mot-vedette n'apparaît pas dans d'autres dictionnaires et il ne se prononce pas comme le furigana. Il doit être corrigé à la main. Cela concerne 17 233 mots-vedette, soit 21 % du total. Il reste encore 12 872 mots-vedette à corriger.
- Le mot-vedette est vide. Il faut le rajouter à la main. Cela a concerné 10 329 mots-vedette, soit 12 % du total.

5. Corrections manuelles

5.1. Mots-vedette

Le remplissage de mots-vedette non-détectés a été effectué de manière automatique avec 65 % de succès. Nous avons utilisé la prononciation en hiragana pour effectuer une recherche dans un autre dictionnaire : le Super Daijirin. Si dans les deux dictionnaires le mot-vedette n'a pas d'homophones, nous avons remplacé le mot-vedette non détecté par celui du Super Daijirin.

Le reste a été effectué manuellement par des contributeurs volontaires, qui sont étudiants français ou japonais, japonisants, chercheurs, linguistes, etc. Sur la page d'accueil du projet, une liste des 20 mots-vedette non détectés classés par fréquence de leur prononciation est affichée pour motiver les contributeurs potentiels. La liste est mise à jour automatiquement tous les soirs. Elle peut également être mise à jour manuellement par un contributeur s'il a corrigé tous les mots-vedettes affichés. Dans la plupart de cas, ce sont des noms anciens de plantes ou d'animaux, des noms d'outils désuets, d'anciens kanji, etc. Par exemple : 薊 (azami, chardon), 攄所 (yoridokoro, fondement).

Le remplissage se fait d'abord par un copier-coller depuis la page PDF scannée du Cesselin vers l'interface d'édition en ligne de Jibiki. En effet, sur MacOs, le lecteur de PDF inclut un outil de reconnaissance optique de caractères. Lorsque le copier-coller échoue à cause de la présence de kanji trop anciens ou d'une mauvaise reconnaissance de caractères, il faut consulter des dictionnaires de kanji, qui permettent de trouver et copier des kanji par des composants (偏 (hen), 旁 (tsukuri)) ou par le nombre de traits (kakusû, 画数)[9] lorsque l'on ne connaît ni leur prononciation ni le sens.

Cette opération a concerné 10 329 mots-vedettes et a duré 2 ans et demi. Elle a été effectuée en grande majorité par 3 contributeurs. Mutsuko en particulier s'est astreinte à des séances de corrections quotidiennes de 30 minutes pour un résultat d'environ 20 mots-vedette. Au total, Mutsuko a corrigé 7 194 mots-vedette.

5.2. Compteurs : quantificateurs et classificateurs

En japonais, il existe des lexèmes qui indiquent la classe des noms, lorsqu'ils apparaissent dans une expression quantitative. Ils dépendent du type de référents ou de l'observation de référents par celui qui parle. Comme il n'existe pas de lexèmes correspondants en français dans le cas général, ils causent souvent des problèmes pour la

9 http://kanji.jitenon.jp/
http://kanjitisiki.com/
https://www.sanseido.biz/main/dictionary/hanrei/daijirin_v3.aspx

traduction automatique français-japonais. Nous avons donc annoté les lexèmes du Cesselin apparaissant comme classificateurs/quantificateurs dans nos listes (Tomokiyo & Boitet, 2016), (Tomokiyo et al., 2017).

Exemples de quantificateurs :

- 家畜３０頭 (kachiku 30 tou, trente têtes de bétail)
- 一枚のＴシャツ (ichimai no ti-shatsu, un T-shirt)
- 山のような問題 (yama no youna mondai, un tas de problèmes)

Exemples de classificateurs :

- 二言、三言の会話 (futakoto, mikoto no kaiwa, des bribes de conversation)
- 皮肉をちくりと (hiniku wo chikurito, une pointe d'ironie)

Cela concerne 253 articles pour l'instant pour la fonction de quantificateur, et il reste encore à peu près 50 articles à vérifier. Quant aux classificateurs, nous avons pris l'approche linguistique en tâtonnant : fabrication de corpus, développement de la liste de KWIC, et constatation des expressions figées. Nous avons rassemblé actuellement 86 lexèmes ayant une fonction de classificateur.

5.3. Classification et annotation des exemples

Le dictionnaire Cesselin contient des exemples divers pour aider à la compréhension des mots. Mais il manque des indications d'usage telles que proverbe, expression figée, classificateur/quantificateur, usage ordinaire, usage dans tel ou tel domaine. Nous avons donc décidé de classer tous les exemples et d'annoter leur usage. Pour ce faire, nous avons fait une expérimentation avec une hypothèse que des expression figées ou des proverbes ne soient pas correctement traduits par la traduction automatique, et par conséquent qu'on puisse distinguer les exemple ordinaires et d'autres.

Nous avons choisi 500 exemples dans le Cesselin, et traduit les exemples japonais et ses traductions françaises en anglais par Google translate[10] pour comparer les deux traductions (Tomokiyo et al. 2018).

Par exemple, pour le verbe 飲む (nomu, boire) :

(a) un proverbe : 飲まぬ酒には酔わぬ[11] (nomanu sake ni wa yowanu, iI n'y a point de fumée sans feu).

(b) une expression figée : 彼は妻君に飲まれている[12] (kare wa saikun ni nomareteiru, sa femme le berne).

(c) un exemple d'usage ordinaire: 水を飲む (mizu wo nomu, boire de l'eau).

La procédure de classification manuelle, qui a été expérimentée pour 500 exemples, est donnée ci-dessous, et nous sommes en train de l'automatiser (Tomokiyo et al. 2018).

10 https://translate.google.com/?sl=en
11 Une traduction mot-à-mot : On ne s'enivre pas de Sake quand on ne l'a pas bu.
12 Une traduction mot-à-mot : Il est avalé par sa femme.

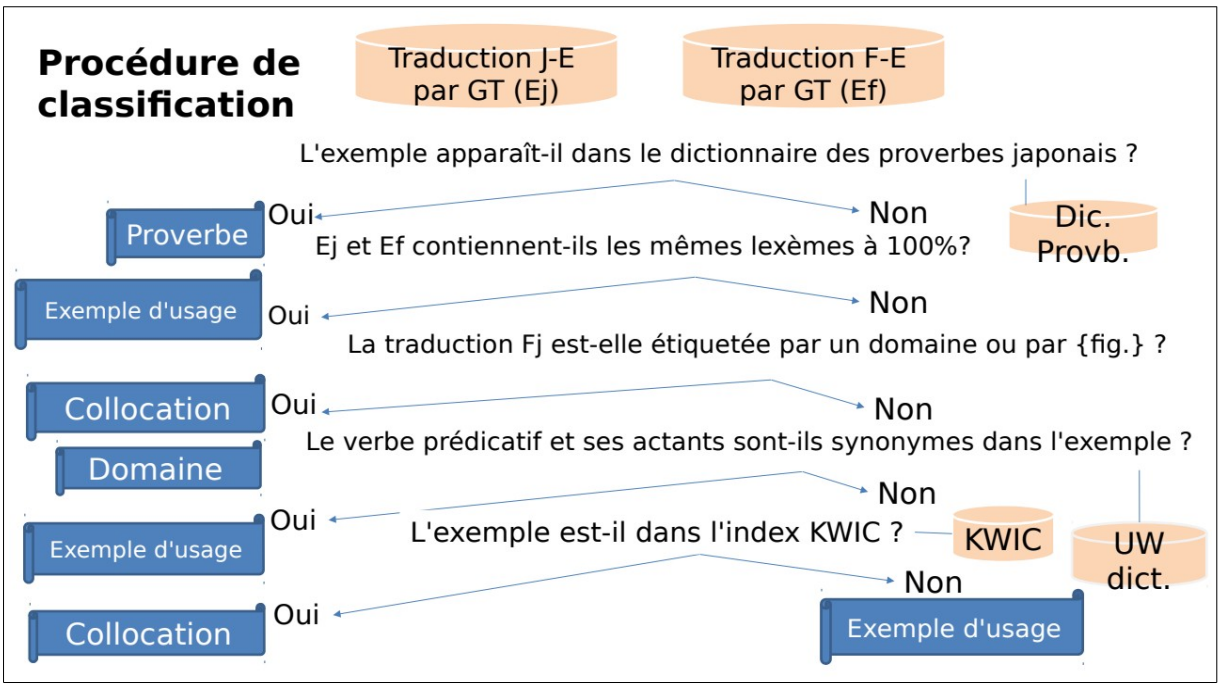

Figure 6 : Algorithme de traitement des exemples

6. Conclusion et perspectives

Après 3 ans et demi de vie du projet, nous pouvons affirmer que l'expérience est concluante. Le dictionnaire est en constante évolution grâce aux corrections manuelles des utilisateurs qui le consultent. L'ensemble de données ainsi constitué sert également de matière première pour concevoir d'autres projets de traitement automatique des langues entre le français et le japonais comme par exemple la lecture active.

Nous avons plusieurs chantiers en perspective pour améliorer de manière automatique les données. La classification des exemples est le plus avancé. Nous envisageons également de traiter les doublons entre les articles du JMdict et les collocations des articles du Cesselin.

Nous envisageons également lorsque nous trouverons le temps de le faire, d'appliquer la même procédure de lecture optique puis correction automatique de données sur un dictionnaire imprimé bilingue monodirectionnel français→japonais, le dictionnaire Raguet-Martin (Raguet et al., 1953).

L'étape suivante est de construire un dictionnaire bilingue bidirectionnel en fusionnant les deux dictionnaire monodirectionnels dont les données seront issues majoritairement du Cesselin et du Raguet-Martin.

Bibliographie

Apel, Ulrich (2002) *WaDokuJT - A Japanese-German Dictionary Database.* Papillon 2002 Seminar, 16–18 July 2002, NII, Tokyo, Japan.

Breen, Jim W. (2004) *JMDict: a Japanese-multilingual dictionary.* In: Coling 2004 workshop on multilingual linguistic resources, Geneva, Switzerland, pp. 71–78.

Cesselin, Gustave (1940) *Dictionnaire japonais-français.* Maruzen, Tokyo, juillet 1940, 2340 p.

Desperrier, Jean-Marc (2002) *Analyze [sic] of the results of a collaborative project for the creation of a Japanese-French dictionary*. In: Proceedings of Papillon 2002 Seminar, 16–18 July 2002, NII, Tokyo, Japan.

Mangeot, Mathieu (2016) *Collaborative Construction of a Good Quality, Broad Coverage and Copyright Free Japanese-French Dictionary.* International Journal of Lexicography, Volume 31, Issue 1, 1 March 2018, Pages 78–112; doi: 10.1093/ijl/ecw035; 35 p.

Mangeot, Mathieu (2015) *Construction collaborative d'un dictionnaire japonais-français de qualité, à large couverture et libre de droits.* Rapport interne LIG, 31 p.

Raguet, Émile ; Martin, Jean-Marie. (1953) *Dictionnaire français-japonais*, Hakusuisha, Tokyo, 1467 p.

Tomokiyo, Mutsuko, Boitet, Christian ; Mangeot, Mathieu (2018) *Towards an Automatic Classification of Illustrative Examples in a Large Japanese-French Dictionary Obtained by OCR.* Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing, pp. 112 - 121., August 2018, Santa Fe, United States.

Tomokiyo, Mutsuko, Mangeot-Nagata, Mathieu ; Boitet, Christian (2017) *Development of a classifiers/quantifiers dictionary towards French-Japanese MT.* MT Summit 2017, 18-22 September 2017, Nagoya, Japan.

Tomokiyo, Mutsuko, Mangeot-Nagata, Mathieu ; Christian Boitet, Christian (2018) *Analyse et classification d'exemples illustratifs dans le dictionnaire "Cesselin" en utilisant Google Traduction et un dictionnaire UNL-UWs.* 11es journées du réseau Lexicologie Terminologie Traduction 2018, Sep 2018, Grenoble, France.

Tomokiyo, Mutsuko ; Boitet, Christian (2016) *Corpus and dictionary development for classifiers/quantifiers towards a French-Japanese machine translation.* 5th Workshop on Cognitive Aspects of the Lexicon CogAlex@COLING 2016, 12 December 2016, Osaka, Japan.