Towards an Automatic Classification of Illustrative Examples in a Large Japanese-French Dictionary Obtained by OCR

Mutsuko Tomokiyo, Christian Boitet, Mathieu Mangeot

Towards an Automatic Classification of Illustrative Examples in a Large Japanese-French Dictionary Obtained by OCR

Mutsuko Tomokiyo
UGA, LIG-GETALP
Grenoble
mutsuko.tomokiyo@imag.fr

Christian Boitet
UGA, LIG-GETALP
Grenoble
christian.boitet@imag.fr

Mathieu Mangeot
UGA, LIG-GETALP
Grenoble
mathieu.mangeot@imag.fr

Abstract

This paper focuses on improving the Cesselin, a large, open source Japanese-French bilingual dictionary digitalized by OCR, available on the web, and contributively improvable online. Labelling its examples (about 226,000) would significantly enhance their usefulness for language learners. Examples are proverbs, idiomatic constructions, normal usage examples, and, for nouns, phrases containing a quantifier. Proverbs are easy to spot, but not the other types. To find a method for automatically or at least semi-automatically annotating them, we have studied many entries, and hypothesized that the degree of lexical similarity between results of MT into a third language might give good cues. To confirm that hypothesis, we sampled 500 examples and used Google Translate to translate into English the Cesslin Japanese expressions and their French translations. The hypothesis holds well, in particular for distinguishing examples of normal usage from idiomatic examples. Finally, we propose a detailed annotation procedure and discuss its future automatization.

1 Introduction

The Cesselin Japanese-French bilingual dictionary was edited by a French missionary, Gustave Cesselin (1873-1944), and published in 1939 and 1957 in Japan (82,703 entries and 2,345 pages)[1]. We have converted it to an electronic form by using an OCR (optical character reader) and made some improvements. An example of the entry for 飲む *nomu* "drink" in original and online form on the Jibiki platform is given in Figure 2 below (before the references).

Our aims are to improve, update and complete the Cesselin to make it available as a modern on-line dictionary for Japanese or French language learners or researchers in Japanese studies, and also to contribute to some improvement in the translation performance of Japanese↔French machine translation (MT) systems. About 60,000 other articles have already been added from other free sources.

The Cesselin includes knowledge on spoken and written Japanese[2], and rich illustrative examples containing the headwords. It is however not satisfactory for modern users, because it contains outdated

[1]A few other on-line Japanese-French dictionaries exist: the Diko (3000 entry words, Diko), the *Dictionnaire français-japo*nais (15,000 entry words, Lexilogos), the *Dictionnaire japonais* (24,000 entry words, Assimil), and the *Dictionnaire Glosbe* (entry words unknown, Glosbe), etc. The number of words in these dictionaries may change, because they are going to be edited and extended in a collaborative way.

[2]The Japanese writing system has been fixed by the Japanese government in 1986, in order for the writing form to conform to pronunciation.

Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing, pages 112–121
Santa Fe, New Mexico, USA, August 20, 2018.

phonetic descriptions, old Chinese characters and old forms of Okurigana.[3] Also, its examples are given without any indication[4] concerning their types. As in many dictionaries, the examples are given to help understanding various meanings and usages of words, and to show their grammatical constructions, and their usages in collocations, proverbs, and quantified phrases. Information on the exact type of each example would help users looking for specific information such as numerical quantifiers, and enable MT developers to methodically deal with lexical ambiguities (Tomokiyo et al., 2016). It would also help language learners or researchers to practice Japanese or French, and to deepen their understanding of the Japanese culture.

In Section 2, we explain an experiment made in order to investigate what kinds of examples are included in the Cesselin. After getting a general impression while correcting the OCR results of about 15,000 entries, we have manually analyzed 500 examples (out of about 226,000), taken from 25 complex entries, established a classification of examples, and proposed classification criteria. In Section 3, we propose an automatable step by step procedure to classify and annotate the remaining 225,500 examples according to these criteria. In Section 4, we briefly present the Universal Networking Language (UNL) Universal Words (UW) dictionary we used to test words in English translations for synonymy. We also examine some technical aspects of the automatization of our procedure. That step, now beginning, will likely save much human expert time, because it will only be needed to "post-edit" automatically generated annotations, and, we hope, to correct less than 5% of them.

2 Case studies on using results of Google Translate[5] for the classification

To find a classification of the examples that would be useful and amenable to automatic or at least semi-automatic annotation, we have studied many entries, and hypothesized that the degree of lexical similarity between results of MT into a third language might give good cues. To confirm that hypothesis, we selected 500 examples from complex entries (having more than 25 examples in average) and used Google Translate (GT) to translate into English both their Japanese expressions and their French translations.[6] We then compared the two translation outputs from the point of view of lexical similarity. Figure 1 illustrates the process.

That comparison confirmed the assumption[7] that, if an example is a proverb or if a word in an example is employed in a collocation, the J→E translated example (Ej) usually contains lexemes that completely differ from those of the F→E MT-translated example (Ef), because in proverbs and collocational expressions including classifiers/quantifiers, words have a tendency to be used in figurative meaning, and GT does not in general produce adequate translation outputs for them, but very often literal and incorrect translations. Thus, the two outputs contain different lexemes,[8] and one can consequently distinguish proverbs and collocations from other usage examples.

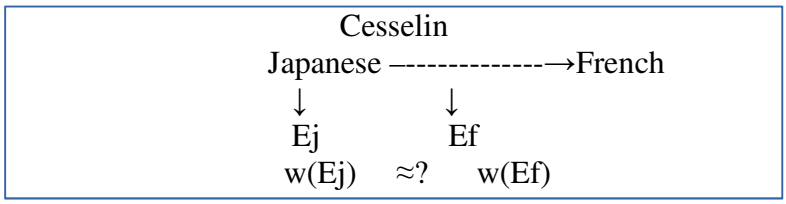

Figure 1: Assumption: according to the similarity of the bag of words w(Ej) and w(Ef), the example is collocational, proverbial, idiomatic, or concerns a quantifier/classifier

[3]A Japanese word is written in Kanji (漢字, Chinese character) and Hiragana (平仮名, Japanese character), in Kanji only, or in Hiragana and/or Katakana (片仮名) only. Okuriganas (送り仮名) are hiragana suffixes following Kanji stems. The rules for Okuriganas changed several times, and were standardized by the Japanese government in 1973.

[4]Excepting indications for figurative meaning usage by the label {fig} and for some domain-specific names.

[5]`https://translate.google.fr/?hl=fr - fr/ja/La démocratie n%27a plus de cote.`

[6]We have the premise underlying that translations into French proposed by the Cesselin are correct.

[7]The assumption is grounded on our studies on Japanese, French and English classifiers and quantifiers (Tomokiyo et al., 2017).

[8]We don't consider the fillers such as *John* and *Mary* in *John pulled Mary's leg*, only *pulled* and *leg*.

2.1 Japanese-English and French-English translation of the 500 examples by GT system

The structure of an entry in the Cesselin is as follows: head word (in Japanese), pronunciation given in two different forms (in Japanese and in transliteration in Latin characters, called ローマ字 (Ro-maji)), conjugated forms for verbs and adjectives, part of speech label, definitions (in French), and examples in Japanese, with their transliteration in ローマ字, and their translation or translations into French. Here is an extract of the entry for the verb *nomu* "drink" from the Cesselin dictionary.

```
nomu[ma, mi, me] 飲む - 呑む【のむ】v.t.
1. Boire, avaler.[9]
2. Aspirer, sucer, fumer.[10]
3. Prendre.[11]
4. 飲まねば薬も効能なし(Nomaneba kusuri mo kônô nashi) "Qui veut
la fin veut les moyens.[12]"
5. 飲まぬ酒には酔はぬ (Nomanu sake ni wa yowanu) "II n'y a point
de fumée sans feu.[13]"
```

We have submitted to GT 500 input examples[14] contained in about 20 complex entries of the Cesselin (Table 1): 飲む *nomu* "drink," 買う *kau* "buy," 走 *hashiru* "ran," 居る *iru* "stay, be," 来る *kuru* "come," 思う *omou* "think," 言う *iu* "say," 可愛い *kawaii* "pretty," 強い *tsuyoi* "strong," 安い *yasui* "cheap," すぐ *sugu* "soon, immediately," から *kara* "because, from, since," 家 *ie* "house, family," 夜 *yoru* "night," 足 *ashi* "leg," 車 *kuruma* "car," 年 *nen* "year," 何時 *nanji* "what time," だけ *dake* "only," ながら *nagara* "while," 必要 *hitsuyou* "necessity," もう *mou* "already," 事 *koto* "thing."

(a)	(b)	(c)	(d)	(e)
Japanese examples in the Cesselin	**GT J→E translations**	**GT F→E translations**	**Cesselin French translations**	**Matching results for J-E and F-E[15]**
水を飲む。(Mizu wo nomu)	Drink water.	Drink water.	Boire de l'eau.	100%
飲まぬ酒には酔わぬ。(Nomanu sake niha yowanu.)	I'm not intoxicated with drinking alcohol.	There is no smoke without fire.	II n'y a point de fumée sans feu.	(Ø P)
飲んだり吐出したりする。(Nondari hakidashitari suru.)	It drinks or spits out.	Swallow and turn in turn.	Avaler et rendre tour à tour.	Ø
飲んだり吐出したりする。(Nondari hakidashitari suru.)	It drinks or spits out.	Accept and refuse.	{fig} Accepter et refuser.	Ø {fig} (collocation)

Table 1: Excerpt of translations obtained by GT (J→E and F→E)

[9]To drink, to swallow
[10]To inhale, to suck, to smoke
[11]To take
[12]He who wills the end wills the means.
[13]There is no smoke without fire.
[14]The chosen words belong to the 200 most frequent words in the frequency list except for 助詞 *joshi* "postpositions":『現代日本語書き言葉均衡コーパス (Gendai Nihongo kakikotoba kinkou ko-pasu』』語彙表 (Giohyou, The Balanced Corpus of Contemporary Written Japanese (BCCWJ) (http://pj.ninjal.ac.jp/corpus_center/bccwj/freq-list.html))
[15]Matching results are judged by a linguist in the experiment.

2.2 Comparison and contrast of the two translations results

We have asked the following questions concerning the two outputs by GT:

- do the two outputs include totally the same lexemes?
- when they include different lexemes, what kind of lexemes are common to both?

Column (e) in Table 1 shows a manual comparison of the J→E translations (Column (b)) outputs with the F→E translations (Column (d)) from the point of view of lexemes in an example.

Table 2 shows the comparison of the two translation outputs.

items	explanation	count	percentage
(1) 100% concordance (100%)	GT J→E and F→E translations of an example include exactly the same words.	63/500	12.6%
(2) 100%* concordance (100%*)	GT J→E and F→E translations of an example include synonym words	79/500	15.8%
(3) Concordance of words with ⊆ or ⊇ (100%⊆ or ⊇)	The lengths of the GT J→E translation and the GT F→E translation differ significantly, but the two include almost the same words.	26/500	5.2%
(4) Concordance by morphosyntactic analysis (100%**)	GT J→E and F→E translations of an example include the same words for the predicate verb and its principal actants.	15/500	3.0%
(5) No-concordance with{fig}label (Ø{fig.})	GT J→E and F→E translations of an example don't include any common words and the French example is marked by the {fig.} label.	8/500	1,6%
(6) No concordance (Ø)	GT J→E and F→E translations of an example don't include any common word except articles.	269/500	53.8%
(7) No concordance and proverbs (Ø P)	GT J→E and F→E translations of an example don't include any common content word and the example appears in a Japanese proverbs dictionary.	32/500	6.4%
(8) No concordance with usage domain (Ø D)	GT J→E and F→E translations of an example having an indication of usage domain don't include any common word.	4/500	0.8%
(9) Zero output (x)	Zero output by GT.	4/500	0.8%

Table 2: Comparison of J-E and F-E translations outputs for the 500 examples: different cases

According to this experiment, the examples in the Cesselin can be classified into 5 classes:

- ordinary usage examples to give grammatical information or a context where headwords are used
- proverbs
- collocational expressions
- classifiers/quantifiers, which depend on signified entities
- domain-specific examples

2.3 Explanations for examples annotation

In the following examples for the verb 飲む *nomu* “drink” in Table 3, (a) is a proverb, (b) is a sentence including the form 飲む in a figurative meaning, and (c) is a sentence that is not a proverb nor a collocation, but ordinary usage. When one compares the lexemes in the J→E and F→E translations by GT,

while (a) and (b) don't contain any common word in their predicative part, in (c), the two translations contain two common words (*drink* and *water*).

That confirms that examples showing ordinary usages have a tendency to be translated as sentences having many words (or synonyms) in common. We thought this phenomenon could be a trump card to classify all examples in the Cesselin, because J→E translations by GT[16] of polylexical expressions are almost always word-for-word translations. We suggest that this comes from a lack of information on the figurative usages of words in GT resources (bilingual dictionary and parallel corpus). Hence, when a word in a Japanese example is used in a figurative meaning depending on the context, the translation result almost never lexically matches the F→E translation result.

	Japanese examples	**GT J-E translations**	**GT F-E translations**	**Cesselin translations of Japanese examples**
(a) Proverb	飲まぬ酒には酔わぬ[17] (Nomanu sake ni-ha yowanu)	I'm not intoxicated with drinking alcohol.	There is no smoke without fire.	Il n'y a point de fumée sans feu.
(b) Collocation	彼は妻君に飲まれている。(Kare ha saikun ni nomare-teiru)	He is being drunk by his wife.	His wife is bowing him.	Sa femme le berne.[18]
(c) Ordinary usage	水を飲む。(Mizu wo nomu)	Drink water.	Drink water.	Boire de l'eau.

Table 3: Examples for a proverb, a collocational expression and an ordinary usage

Among ordinary usage examples, proverbs, collocational expressions, classifier/quantifiers and domain-specific examples, we have distinguished proverbs from other types of examples by using a proverb dictionary. Using also labels such as {fig} (in Table 4) and {botanic, zoology….} which are attached to some examples in the Cesselin, we have been able to distinguish collocational expressions and domain-specific examples from other examples, respectively.

3 Examples annotation procedure

We propose the following procedure for classification and annotation of the examples in the Cesselin.

Step 1) Differentiation of proverbs from others

We distinguish proverbs from other examples, using the proverb dictionary.[19]

When Japanese examples appear in this Japanese proverbs dictionary, they are annotated as {`prov`} for *proverb*.

Step 2) Differentiation of ordinary usage examples from others

When lexemes in two translation outputs have 100 percentage of concordance, the example is annotated as {`ordusg`} for *ordinary usage*.

Step 3) Differentiation of collocational expressions and expressions in specific domains

In the Cesselin, the label {`fig`} indicates a figurative usage of headwords, but is not used consistently. When the translation for an example is marked with that label (see {`fig`} in Table 4), we annotate it as {`colexp`} for *collocations*[20], and when the French translation is labelled by {`botanic, zoology`, etc.}, we annotate it as {`domexp`} for *domain-specific usage*.

[16]Of course, translation performance of commercial systems depends on the considered language pair.

[17]A word-to-word translation would be: *One doesn't get drunk on Sake which one doesn't drink.*

[18]In modern French, one would say *Sa femme le trompe* — another illustration of the need to update the content.

[19]新明解故事ことわざ辞典 (Sinmeikai koji-kotowaza jiten, Dictionary of legends and proverbs), 三省堂編修所 (Sanseido Editions), Tokyo. That dictionary contains 7,300 items.

[20]Note that phrases or sentences which include words used in figurative meaning are not always collocational expressions.

items	Japanese examples	GT **J-E** translations	GT **F-E** translations	**French translation of Japanese examples**
Collocation without indication	人を飲んで掛かる(Hito wo non de kakaru)[21]	I'm drinking people and hanging.	Miss someone, look down.	Manquer à quelqu'un, regarder de haut.
Collocation with indication {fig}	飲んだり吐出したりする[22] (Nondari hakidashitari suru)	To drink or to spit.	Accept and refuse.	{fig} Accepter et refuser.

Table 4: Examples of words having a figurative meaning, with and without the label {fig}

Step 4) Differentiation of collocational expressions without any label in Cesselin

There are cases where French translations in Cesselin have no label {fig}, but are used in figurative meaning (Table 5). When their English translations haven't any common lexeme (w(Ej) ∩ w(Ef) = ∅), we classify them as collocational expressions, after referring to a KWIC (Keywords in context) list[23] obtained by using the Sketch Engine[24] software. When an example appears many times in the KWIC list, it is considered as a collocation or a classifier/quantifier (Alda, 2011) and we annotate it as {`colexp`} for *collocational usage*. We compare with the KWIC list, and, if the example is a noun phrase of type "number + noun," or "noun + のような (no-youna, like) + noun,"[25] we annotate it as {`quant`} for *classifiers/quantifiers* (Tomokiyo et al., 2017; Miyagawa, 1989).

headword	**Japanese example**	GT **J→E** translation	GT **F→E** translation	**French translation for the Japanese example**
足 (ashi, foot)	足を洗う[26] (ashi wo arau)	Wash your feet.	Rise from a lower class.	S'élever d'une classe inférieure.

Table 5: Examples of collocational expressions without any label in the Cesselin

Step 5) Checking on synonymy relationship between the main words in the two translated examples

In Table 6, the J→E translation of the example 必要費 *hitsuyou hi* "necessary cost" is different from the F→E translation of *cost* and *expenses*, and the J→E translation of the example 車に乗る *kuruma ni noru* "to get on a car" is different from the F→E translation for *ride* and *get in*, which are the predicative verbs of the sentences. In these cases, we check whether a synonymy relationship holds between the two words by using a (UNL) UW dictionary (see 4.1). If the two words (rather, word senses) are synonymous, the two translated examples are considered to have 100% concordance, and the example is annotated as {`ordusg`} for *ordinary usage*.

[21]To catch the person, drinking him.

[22]The concrete meaning of 飲んだり吐出したりする : *Drinking and spitting alternately*.

[23]The input is a corpus developed by Mathieu Mangeot (Mathieu's corpus) in the framework of the Jibiki project since 2014 (`http://jibiki.fr`). It is a Japanese-French parallel corpus coming from newspapers, novels, the Bible, etc., and contains 9268785 words at the moment.

[24]The Sketch Engine is a corpus management tool, containing 400 ready-to-use corpora. We have added Mathieu's ccorpus to it. See `https://www.sketchengine.co.uk`

[25]E.g. 山のような問題 *yama no youna mondai* "problems like mountain" → a lot of problems

[26]The concrete meaning of 足を洗う is: *Wash one's feet*. By the way, the expression 足を洗う makes sense in concrete as well as figurative usage. We have not yet solved this problem.

headword	Japanese examples	GT J→E translations	GT F→E translations	French translations	comparison
必要 (hitsuyou)[27]	必要費 (hitsuyou hi)	Necessary cost	Necessary expenses.	Dépenses nécessaires.	100%*[28]
車 (kuruma)[29]	車に乗る (kuruma ni noru)	I ride a car.	Get in the car	Monter en voiture.	100%*

Table 6: Two translated examples having synonymous lexemes

Step 6) Analyzing translated examples having different lengths

Examples translated into English have different lengths because of the nominalization[30] of a predicative verb, like (a) in Table 7, as well as difference of registers (formal vs. informal setting, degree of politeness), as in utterances (b) in Table 7, etc.

In example (b), *Excuse me*, *are not you*, and *Mr. Yamada* are common to the two translated examples, and only *Please* and *but* are added in the F→E translation. In this case, the example is syntactically analyzed, and if the predicative verb and its main actants are the same, the two examples are considered to have 100% concordance, and we annotate it as {`ordusg`} for *ordinary usage*.

Expression in the Cesselin	Japanese examples	GT J→E translations	GT F→E translations	French translation for the Japanese examples
a) もう (mou) (J ⊇ F)	もう休みましょう (Mou yasumi masyou)[31]	Let's have a rest now.	Let's rest now!	Reposons-nous maintenant!
b) ながら (nagara) (J ⊆ F)	失礼ながら山田様ではございませんか (Shitsurei nagara Yamada san deha gozaimasen ka)[32]	Excuse me, are not you, Mr. Yamada?	Please excuse me, but are not you Mr. Yamada?	Veuillez bien m'excuser, mais n'êtes-vous pas monsieur Yamada?

Table 7: Two translated examples having different lengths

4 Towards automatic classification

4.1 Usage of a UW dictionary as a pivot of interlingual lexemes

In order to get information on the synonymy relationship between English words in the two translated examples, and also morphosyntactic information, we have used a UW dictionary[33] made available by the UNL project,[34] as a pivot of "interlingual lexemes."

[27] *hitsuyou* = "necessary"

[28] The symbol 100 %* stands for the matching of 100 % of the two translation outputs with synonymy information.

[29] A car

[30] We are not yet engaged in this case.

[31] Let's take a rest now.

[32] Excuse me, are not you Mr.Yamada?

[33] The UNL-UW dictionary contains at the moment 126,9421 headwords for Japanese, 520,305 headwords for French and 1,458,686 headwords for English. The semantic attributes consist of 58 labels and 39 semantic relation labels.

[34] The UNL (Universal Networking Language) project was launched at the Institute of Advanced Studies (IAS) of the United Nations University in Tokyo (UNU) in April 1996 under the aegis of the United Nations University in Tokyo, with financial

A Universal Word (UW) is a character-string which represents a sense of a word. It is made of a headword (usually an English word or term) followed by a list of semantic restrictions between parentheses. A UNL semantic representation of an utterance is a hypergraph, where nodes are UWs having semantic attributes, and arcs bear semantic relations between two nodes or scopes. One node has to bear the `@entry` attribute. A scope is an arc-connected subgraph having an entry node and may be referred to as origin or extremity of an arc.

Here are some examples of entries in the UW dictionary:

(a) `expense(icl>cost)` [`expense` is "included" in `cost`, hence `cost` has its nominal sense here]

(b) `look(agt>thing, equ>search, icl>examine(icl>do, obj>thing))`
[here:
- `icl` in (a) gives synonym information (`examine`)
- `agt` and `obj` in (b) denote the nature of the 2 main actants of this predicative verb (in discourse)
- `equ` in (b) expresses the fact that `look` (maybe the headword should be `look_for`) is linked to `search` by an `equ` path in the UNL ontology map[35] (Uchida et al., 2006).]

4.2 Mechanisms needed for automatic annotation

Before starting this research, we produced a morphosyntactic analysis of the Cesselin examples by using MeCab (Mangeot 2016).[36] Hence, the Japanese examples are already segmented in words and annotated according to a grammar based on Hashimoto's grammar.[37]

In order to build an automatic classifier/annotator of the Cesselin examples, we further need:

(1) a mechanism to match examples in the Cesselin with proverbs contained in one or more proverb dictionaries. As no Japanese proverb dictionary is available in electronic form (at least .epub), a first challenge is to build an online Japanese proverb database. We envisage to use the same method as that used for computerizing the Cesselin.

(2) a mechanism to call GT on an example and its translation, and to evaluate the lexical similarity of the two translation outputs. The first part is easy, while the second needs the three following resources.

(3) a mechanism to call an English morphosyntactic analyzer on the translated examples, to get the lemmas and possibly other attributes (e.g., number, determination). PhpMorphy would be a possibility (`http://phpmorphy.sourceforge.net/dokuwiki/`).

(4) a mechanism to check the synonymy between two words by comparing the UWs having their lemmas as headwords in the UW dictionary. For this, we intend to put the UW dictionary in the form of a lexical network (it has been made with an earlier version using the PIVAX/Jibiki lexical database).

(5) a mechanism to access a KWIC list produced from the set of Japanese proverbs.

5 Conclusion and perspectives

Labelling the examples (about 226,000) of the Cesselin would significantly enhance their usefulness for language learners. We have hypothesized that the degree of lexical similarity between results of MT (by GT) into English might give good cues. That hypothesis has been confirmed by manually applying a procedure derived from it on 500 examples, and getting 100% correct annotations. Incidentally, that success confirms that handling of polylexical expressions by MT systems (even neural ones) is still very

---

support from the ASCII corporation and IAS. See `http://www.undl.org/unlsys/unl/unl2005/attribute.htm` (Uchida et al., 2006).

[35]The semantic relation labels are created from UNL ontology, which stores all relational information in a lattice structure, where UWs are interconnected through relations including hierarchical relations (10 levels) such as `icl` (a-kind-of) and `iof` (an-instance-of), and mean headword's sub-meaning, respectively. See `http://www.undl.org/unlexp/`.

[36]Spoken Japanese has no space between words in a sentence.

[37]Hashimoto's grammar was introduced by the linguist Shinkichi Hashimoto (1946), and is widely adopted in the educational field in Japan.

poor. We hope our work could help improve them, at least for the J-F pair. We are working towards automating that procedure and applying it to the remaining 265,500 examples. Of course, we don't hope to get 100% correct annotations on this large set, but we hope that corrections done by cooperative online post-editing, will be limited to 5% or 10%.

We are also planning to study whether that method could spot corresponding proverbs, collocational expressions and quantified expressions in bilingual aligned J-F corpora. It will be a good surprise if it works, because sentences in usual texts tend to be quite longer than dictionary examples, which are in general not very long. Also, they rarely include proverbs, which are frequent in dictionary examples, and easy to spot. However, our method of differentiating collocational expressions from other types of expressions might also work in general documents. We plan to test this in the future. Also, it might be possible to disambiguate polylexical expressions by describing the context where a word appears in the UW dictionary (Uchida et al., 2006).

Figure 2: The *nomu* "drink" entry in the original Cesselin and the online Cesselin/Jibiki (18 examples).

References

Mathieu Mangeot. 2016. Collaborative construction of a good quality, broad coverage and copyright free Japanese-French dictionary. HAL-01294566.

Alda Mari. 2011. *Quantificateurs polysémiques.* Université Paris-Sorbonne, Vol.23, France.

Sigeru Miyagawa. 1989. *Structure and case marking in Japanese.* Syntax and Semantics, Vol. 22, New York.

Mutsuko Tomokiyo, Mathieu Mangeot, and Christian Boitet. 2016. Corpus and dictionary development for classifiers/quantifiers towards a French-Japanese machine translation, COLING, CogALex 2016, pages 185-192, Japan.

Mutsuko Tomokiyo, Mathieu Mangeot, and Christian Boitet. 2017. Development of a classifiers/quantifiers dictionary towards French-Japanese MT, MT-Summit 2017, Research Track, Vol 1, pages 216-226, Japan.

Hiroshi Uchida, Meihyin Zhu, and Tarcisio Della Senta. 2006. *Universal Networking Language*. UNDL Foundation, Japan.