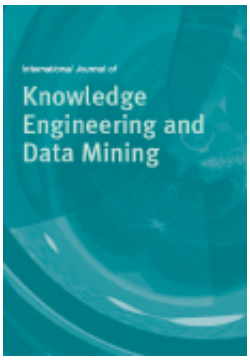

International Journal of Knowledge Engineering and Data Mining

ISSN online: 1755-2095 - ISSN print: 1755-2087
https://www.inderscience.com/ijkedm

A novel practical algorithm for strong and weak synonyms extraction with simple equality operation of web operational machine translation systems results

Mawloud Mosbah

DOI: 10.1504/IJKEDM.2022.10049937

Article History:

Received:	14 March 2022
Last revised:	16 August 2022
Accepted:	17 July 2022
Published online:	10 October 2022

A novel practical algorithm for strong and weak synonyms extraction with simple equality operation of web operational machine translation systems results

Mawloud Mosbah

LRES Laboratory,
Informatics Department,
Faculty of Sciences,
University 20 Août 1955,
Skikda, Algeria
Email: mos_nasa@hotmail.fr

Abstract: In this paper, we address synonyms extraction through a simple algorithm based on equality operation of translated segments generated by four free of charge web machine translation systems namely: *Google Translate*, *Promt*, *Babylon* and *Bing*. The synonyms are classified into two kinds: strong synonyms with different vocabularies, and weak synonyms with the same vocabularies and different part of speech nature. The obtained results, using a dataset of ten theses abstracts, published into the web, seem to be very promising. Moreover, according to this study, it is advised that weak synonyms should receive more attention especially into synonyms substitution for machine translation evaluation.

Keywords: synonyms extraction; weak synonyms; strong synonyms; machine translation evaluation.

Reference to this paper should be made as follows: Mosbah, M. (2022) 'A novel practical algorithm for strong and weak synonyms extraction with simple equality operation of web operational machine translation systems results', *Int. J. Knowledge Engineering and Data Mining*, Vol. 7, Nos. 3/4, pp.271–285.

Biographical notes: Mawloud Mosbah is an Assistant Professor at the Informatics Department, University 20 Août 1955 of Skikda and member of LRES Research Laboratory within the same university. He has received his Doctor of Sciences in Informatics from the University 20 Août 1955 in 2017. His main areas of interest and expertise include information retrieval, image retrieval and automatic natural language processing.

1 Introduction

Machine translation, as an automatic natural language processing application, consists of making the passage, for a segment of text or voice, from a source language to a target language trying to keep the same meaning. Some pre-processing steps have to be considered by machine translation evaluation before applying the assessment metrics

such as synonyms substitution, lemmatisation and stemming. The synonyms substitution pre-treatment phase aims to extend the reference segment through giving the different eventual equivalences may have more eventual intersections with the target segment to be evaluated. Consulting of the literature reveals that there is a lot of spectrum of approaches for synonyms extraction. The major of them generate synonyms as unigrams or single words qualified as strong synonyms with different vocabularies.

In this paper, we aim to address synonyms extraction differently through considering phrases or n-grams rather than single words. The introduced solution is based on machine translation task. Indeed, the solution utilises the segments translated by four free of charge web machine translation systems namely: *Google Translate*, *Promt*, *Babylon* and *Bing*. In addition to the classical synonyms with different vocabularies, known here as strong synonyms,, some phrases to be focused on, known as weak synonyms, share the same morphological words but with different part of speech (POS) natures and in consequence with different grammar rules.

The remainder of the paper is organised as follows: in Section 2, we present different synonyms extraction approaches. We present our proposal in Section 3. Section 4 depicts some results and discussion. We conclude the paper in Section 5.

2 Synonyms extraction

Synonymy relation is one of the lexical semantic relations dealing with words of different forms having the same or similar meaning. According to Naser-Karajah et al. (2021), two words are considered synonyms if they can be interchanged in the same sentence without changing its meaning. In Miller et al. (1990), authors considered that two expressions are synonymous if the substitution of one for the other never changes the truth value of a sentence in which the substitution is made. Synonyms substitution is very important in many applications of automatic language processing such as information retrieval (Kumari et al., 2016) especially in query expansion (Abbache et al., 2018), linguistic steganography (Xiang et al., 2018), entity detection (Ananthanarayanan et al., 2008; Wang et al., 2020), question-answering systems (Esposito et al., 2020), text summarisation (Yadav and Meena, 2016), and machine translation evaluation through extending the reference segment to be compared with the machine translation output fragment.

Synonyms extraction (called also synonym acquiring or synonym detection), as a natural language processing task, proceeds to identify and detect lexical items having identical or similar meanings in some contexts. Unfortunately, as we know, there is only one recent survey for synonyms extraction (Naser-Karajah et al., 2021) although there is a spectrum of approaches namely: rule-based approach (Kazi and Kazi, 2019; Al-Mathan and Al-Khalifa, 2021), machine learning (Mohammed, 2020; Murata et al., 2014), word embeddings (Al-Mathan and Al-Khalifa, 2021; Mohammed, 2020), statistics (Inkpen, 2017; Ababneh et al., 2021; Takeuchi, 2008), translation mirror approach (Muller and Langlais, 2011), and lexico-syntactic like distributional similarity (Lin, 1998). Each approach has its advantages and disadvantages and it is clear that there is no approach which outperforms evidently the others; statistics approach is time consuming while machine learning and some techniques tied to word embedding approach are also time consuming and they require a learning database. Rule-based and lexico-syntactic approaches are tied to the addressed language and the generalisation for other languages

is not possible. For translation mirror approach, it is clear that it has the same disadvantages of translation and its performance is tied strictly to the performance of the considered translators. In Kazi and Kazi (2019), authors have adopted rule-based approach to detect automatically synonyms between a considered ontology and a conceptual data model (*CDM*). The rules are not linguistic but they are structural rules related to the structure of the considered resources namely the ontology and the *CDM*. An illustrative example has been detailed and the value of 38% has been yielded as the highest ratio of synonyms elements. In Al-Mathan and Al-Khalifa (2021), authors have presented a novel solution to filter synonyms extraction by proposing the SynoExtractor pipeline used to filter similar word embeddings to retain synonyms based on specified linguistic rules. The solution, trained with *CBOW* and *SG* models, conducted on *King Saud University Corpus of Classical Arabic* (*KSUCCA*) and *Gigaword* corpus, have yielded respectively 21% and 25% improvements. It outperforms the *sketch engine thesaurus* for synonym extraction by 32% in terms of *mean average precision* (*MAP*). Although the solution has addressed Arabic language, authors claimed that it may be generalised no matter what the considered language. In Mohammed (2020), authors have considered word embedding, using word2vec, as an input to train a feed-forward neural network in order to distinguish between synonyms and other semantically related words. The experiments conducted on *SimLex-999* dataset extended with *WordNet* yielded 0.76 as the best setting accuracy. Authors, in Murata et al. (2014), have conducted automatic selection of an appropriate synonym among multiple candidate synonyms for a certain sentence using machine learning. They claimed 0.86 as a performance with 0.16 of enhancement regarding the baseline method. In addition, they classified synonym pairs into two categories: those that tend to require proper synonym selection for a certain sentence and those that do not require proper synonym selection. In Inkpen (2007), author used the web as a corpus to compute scores based on mutual information. In Ababneh et al. (2021), authors have presented a new efficient model in synonyms extraction called *noun-based distinctive verbs* (*NBDV*) where *tf-idf* weighting scheme is replaced by a novel weighting scheme called the *orbit weighing scheme* (*OWS*). In Takeuchi (2008), author has described that a graph-based co-clustering approach is suitable for extraction of verb synonyms from large-scale text. The proposed bipartite graph algorithm can produce clusters of verb synonyms as well as noun synonyms taking into account word co-occurrence between the verb and its argument. In Muller and Langlais (2011), authors have evaluated, experimentally, mirror translated similarities, which defines similar terms as terms that often share their translation according to multilingual corpora, compared to the more widespread distributional approach, based on some distributional regularities such as co-occurrence and syntactic associations. According to the experiments conducted on *Wacky* corpus, they concluded that mirror translations provide a better filter, with simple computation, than syntactic distribution similarity. In Lin (1998), author has presented a new evaluation methodology for the automatically constructed thesaurus, based on the distribution patterns of terms. Compared to two manual thesauri (*WordNet* and *Roget*), the evaluation results show that the generated thesaurus is significantly closer to *WordNet* than *Roget* thesaurus is.

To note that many resources have been used by the community of researchers for detecting, acquiring, and extracting of synonyms such as dictionaries (Wang and Hirst, 2009; Muller et al., 2006; Blondel and Senellart, 2002; AlMaayah et al., 2016), WordNet (Lombardi and Marani, 2015), web of data (Inkpen, 2007; Lombardi and Marani, 2015), Wikipedia (Bøhn and Nørvåg, 2010; Simanosky and Ulanov, 2011), corpus (Takeuchi, 2008) and encyclopaedia (Lu et al., 2009).

There are two categories for synonyms (Al Ameed et al., 2006): absolute synonyms, having meanings identical in all respects and in all contexts, and partial synonyms, having meanings identical in only some contexts. In this paper, we address synonym detection issue through dealing with other two synonym categories, namely: strong synonymy for synonyms with different vocabularies and weak synonymy for synonyms with the same stem but with different POS. To the best of our knowledge, there is no work which addresses weak synonyms, all the works of the literature, in the context of synonyms detection, deal with strong synonymy.

3 Our synonyms extraction solution

Our proposed solution, for extraction of synonyms, relies on the results returned by four free web machine translation systems, namely: *Google Translate*, *Promt*, *Babylon* and *Bing*. The proposed solution is based then on translation mirror approach (Muller and Langlais, 2011) with the idea that the phrases having the same interpretation are qualified as synonyms. It is worthy to note that the obtained synonyms belong to the target language rather than the source one. As depicted in Figure 1, the execution scenario of the system is then as follows: the user submits a segment of text to be translated from a source language to a desired target language. This segment is communicated to the four considered free web machine translation systems that answer with four associated segments, in the target language, considered as input for *synonyms extraction module*. Although the synonyms extraction module, as presented in its pseudo code of Figure 4, has the same idea based on equality operation between word instances, it distinguishes two types of synonyms: *strong synonyms* where there are completely different vocabularies and *weak synonyms* based on the same stems of vocabularies with different POS. With the detection of weak synonyms, our solution belongs then to lexico-syntactic approach besides the translation mirror approach character. To note that *POS* nature indicates how the word functions in meaning, as well as grammatically, within the sentence. Commonly, there are eight parts of speech: *noun*, *pronoun*, *verb*, *adjective*, *adverb*, *preposition*, *conjunction* and *interjection*. Figure 2 and Figure 3 give respective examples of strong and weak synonyms in English language. For French language, how to detect synonyms is the same as in the case of English with the adequate setting of TreeTagger software (using its French dictionary file).

Figure 1 Our proposal architecture

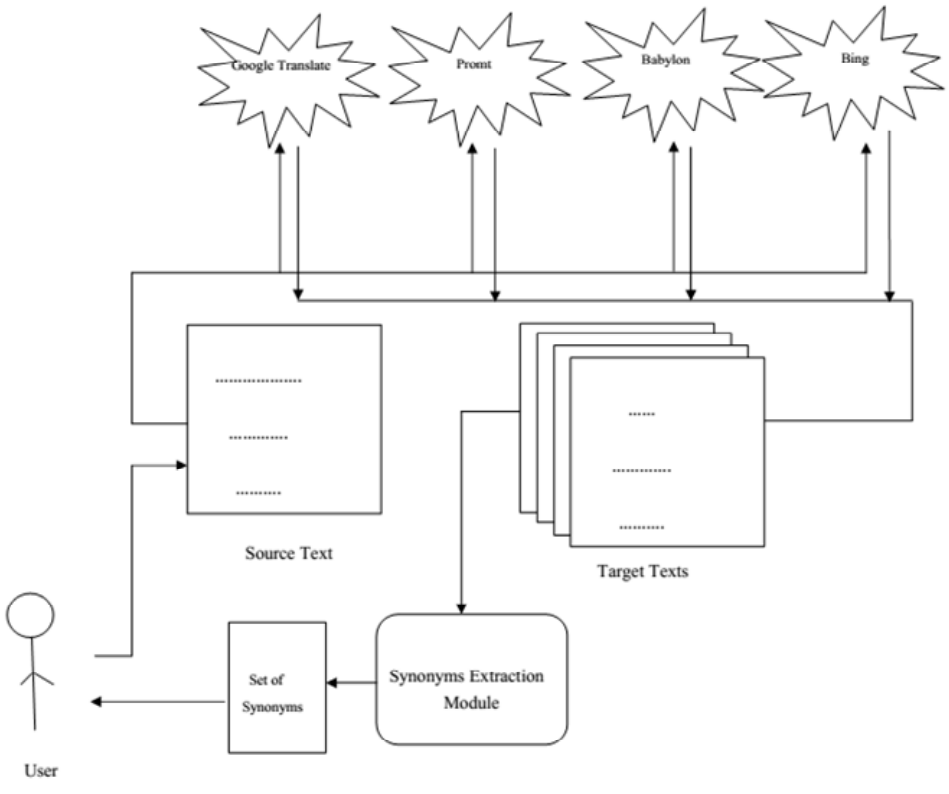

Figure 2 Illustrative English example for strong synonyms

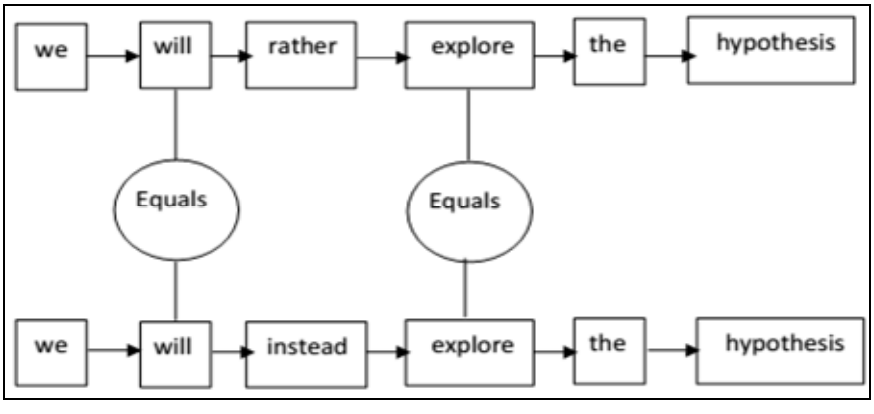

Figure 3 Illustrative English example for weak synonym based on POS and its associative lexical rule

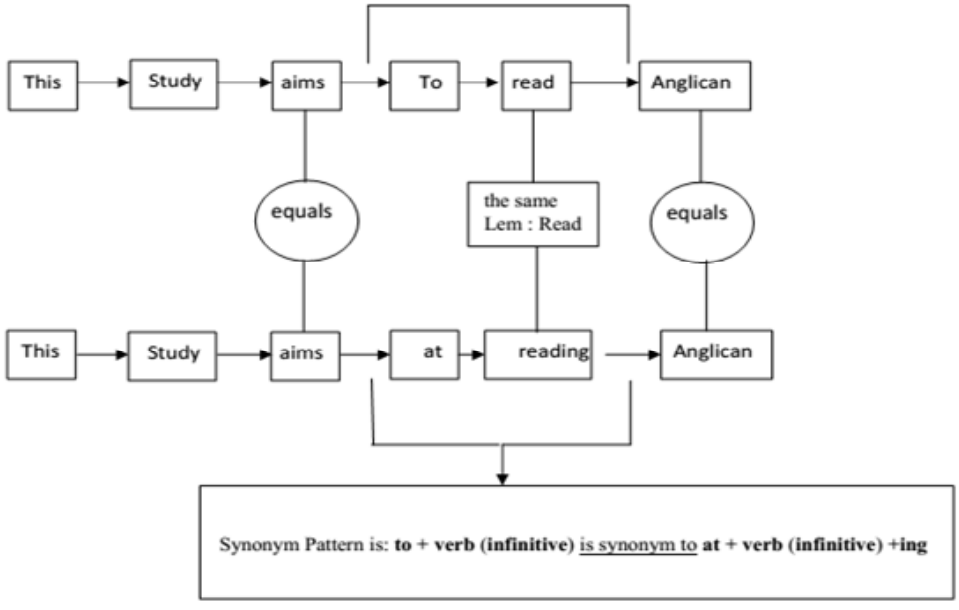

Figure 4 The pseudo code of our algorithm for extraction strong and weak synonyms based on POS

Input: translated sentences generated by a set of free of charge web machine translation systems.

Output: a set of strong and weak synonyms.

For each two translated sentences TS1 and TS2

Explore TS1 tokens and TS2 tokens searching for equal tokens

Consider the words or consecutive words between two equal words as synonyms SN1 and SN2

If there are some words shared by SN1 and SN2 having the same lems or stems then

- We have weak synonyms with lexical rule to be generated.

Else

- We have strong synonyms.

4 Experimental results and discussion

4.1 Dataset and tools

We utilise, as references, ten web published doctoral theses with English and French abstracts given by their authors who are researchers having commonly high level into both English and French and able to well express the problematic of their dissertation. The characteristics of these used theses as a dataset are given in Table 1.

Table 1 Dataset characteristics

Dataset	*# of words in abstract*	
	In English	*In French*
Thesis #1	264	294
Thesis #2	295	293
Thesis #3	289	372
Thesis #4	317	439
Thesis #5	250	275
Thesis #6	578	767
Thesis #7	170	188
Thesis #8	275	313
Thesis #9	287	312
Thesis #10	301	379
Total	3,026	3,632
# in words of detected synonyms	327	301

For web machine translators, we employ four free web machine translation systems namely: *Google Translate* (http://translate.google.com), *Promt* (https://www.online-translator.com/traduction), *Babylon* (https://www.babylon-software.com/?lang=fr) and *Bing* (https://www.bing.com/translator). In order to recognise the POS of the word, we use TreeTagger free software (http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/).

4.2 Experimental results

The experimental results achieved, for strong synonyms, are depicted in Table 2 and Table 3 while some examples of lexical rules dealing with weak synonyms are given later.

4.2.1 Strong synonyms and their experimental results

Table 2 French and English detected synonyms

Thesis #	*French synonyms*		*English synonyms*	
	Synonyms accuracy	*Number of synonyms for the four considered machine translators*	*Synonyms accuracy*	*Number of synonyms for the four considered machine translators*
Thesis 1	0.923	13	0.933	15
Thesis 2	0.894	19	0.904	23
Thesis 3	0.939	33	0.761	22
Thesis 4	0.84	50	0.95	40
Thesis 5	0.769	26	0.956	23
Thesis 6	0.90	40	0.857	49
Thesis 7	0.869	23	0.6	10
Thesis 8	0.883	43	0.909	33
Thesis 9	0.84	50	0.837	37
Thesis 10	0.911	34	0.864	37
Total	0.876	331	0.8571	289

Table 3 Examples of French and English synonyms not based on lexical rules with POS

Examples of French strong synonyms	*Example of English strong synonyms*
‘Mettre l’accent sur’ and ‘souligner’	‘To’ and ‘in order to’
‘Le rapport’ and ‘la relation’	‘To safeguard free’ and ‘to safeguard the gratuity of’
‘Perspicacité’ and ‘point de vue’	‘Issue’ and ‘question’
‘Basé sur’ and ‘fondé sur’	‘Perspicacity’ and ‘insight’
‘La prétendue thèse’ and ‘la thèse dite’	‘to emphasize’ and ‘to underline’
‘De part en part’ and ‘à travers et à travers’	‘Through’ and ‘across’
‘Viser à ‘ and ‘avoir objet pour’	‘Point of view’ and ‘insight’
‘Dans le but de’ and ‘en vue de’	‘As early as’ and ‘from’
‘En vue de’ and ‘pour’	‘To concentrate’ and ‘to focus on’
‘Au niveau de’ and ‘sur le plan’	‘Finely’ and ‘in detail’
‘Étendre’ and ‘élargir’	‘As well as’ and ‘and’
‘La mise en rapport’ and ‘le lien’	‘That’ and ‘which
‘Tantôt’ and ‘parfois’	‘Unicellular’ and ‘single-celled’
‘Cependant’ and toutefois’	‘Adaptation’ and ‘coping’
‘Donc’ and ‘ainsi’	‘Importance’ and ‘key role’
‘Outre’ and ‘en plus de’	‘Thus’ and ‘therefore’
‘La mise au point’ and ‘le développement’	‘Amount’ and ‘quality’
‘Finement’ and ‘en détail’	‘The arrival’ and ‘the advent’
‘Rester’ and ‘demeurer’	‘Beforehand’ and ‘previously’
‘Dès le’ and ‘à partir de’	‘To look at’ and ‘to focus on’

Table 3 Examples of French and English synonyms not based on lexical rules with POS (continued)

Examples of French strong synonyms	*Example of English strong synonyms*
'Au cours de' and 'pendant'	'Unique' and 'particular'
'Notamment' and 'particulièrement'	'To examine' and 'to look at'
'Jeux de données' and 'ensemble de données'	'To seek' and 'to aim at'
'Être basé sur' and 'reposer sur'	'To aim to' and 'to be designed to'
'La sortie' and 'la production'	'Instead' and 'rather'
'Susceptible' and 'qui peut'	'The output' and 'the result'
'Changer' and 'modifier'	
'Spécifique' and 'toute particulière'	
'concrètement' and 'réellement'	

4.2.2 Weak synonyms based on POS with their different lexical rules and an illustrative example

1 Making the passage from 'the + noun1 of noun2' to 'noun2 + 's + noun1' in English.

English
'The grace of God' and 'God's grace'

2 Making the passage from 'adjective + noun' to 'noun + with + noun of the adjective' in English.

English
'Noisy environment' and 'environment with noise'

3 Making the passage from 'one + adjective + noun' to 'a/an + adjective + noun' in English

English
'One important point' and 'an important point'

4 Making the passage from 'to be + adjective + in + noun' to 'have +a/an + adjective + noun' in English.

English
'It is supernatural in character' to 'it has a supernatural character'

5 Making the passage from 'être + past participle' to 'se + verb' in French.

French
'Qui est calculé' and 'qui se calcule'

6 Making the passage from passive voice to active voice and vice versa.

English	*French*
'the method has been evaluated' and 'we evaluated the method'	'La méthode a été évaluée' and 'on a évalué la méthode'

7 Being tolerated with conjugation.

English	*French*
'We evaluated the method' and 'we have evaluated the method'	'On évaluait cette méthode' and 'on a évalué cette méthode'

8 Making the passage from 'adjective' to 'of + noun' in English and from 'noun + adjective' to 'noun + des + noun in plural' in French

English	*French*
'Vocal' and 'of voice'	'traitement linguistique' and 'traitement du langage'
'Doctoral mission' and 'the mission of doctors'	'un parcours doctoral' and 'parcours des docteurs'

9 Making the passage from 'in + adjective + way/manner' to 'adverb' in English and from 'de façon/manière + adjective' to 'adverb' in French.

English	*French*
'In regular way/manner' and 'regularly'	'De façon/manière régulière' and 'régulièrement'

10 Making the passage from 'more + adjective' to 'with more + noun' in English and from 'le plus + adjective' to 'avec plus de + noun' in French.

English	*French*
'More informative' and 'with more information'	'Le plus informatif' and 'avec plus d'information'

11 Making the passage from 'present participle + with' to 'in noun + noun' in English and from 'participe present + avec' to 'en + noun + avec' in French.

English	*French*
'Interacting with' and 'in interaction with'	'Interagissant avec' and 'en interaction avec

12 Making the passage from 'to be more + adjective' and 'adverb' in English and 'pour être bien + adjective' to 'adverb' in French.

English	*French*
'To be more precise' and 'precisely'	'Pour être bien précis' and 'précisément'

13 Making the passage from ‘for both + noun1 + noun2’ and ‘for noun1 as for noun2’ in English and from ‘pour + noun1 + noun2’ to ‘pour + noun1 + que + pour + noun2’ in French.

English	*French*
‘For both performance and efficiency’ and ‘for performance as for efficiency’	‘Pour la performance et l’efficacité’ and ‘pour la performance que pour l’efficacité’

14 Making the passage from ‘proceed/establish/perform + noun’ to ‘verb’ in English and from ‘procéder à + noun + de’ to ‘verb’ in French.

English	*French*
‘I proceed/establish/perform the translation’ and ‘I translate’	‘Je procède à la traduction de’ and ‘je traduis’

15 Making the passage from ‘who/which + verb’ to ‘present participle’ in English and from ‘qui + verb’ to ‘present participle’ in French.

English	*French*
‘Who/which interacts’ and ‘interacting’	‘Qui interagit’ and ‘interagissant’

16 Making the passage from ‘to + verb’ and ‘for + present participle in English and from ‘pour/afin de + verb’ to ‘pour + article + noun + de’ in French.

English	*French*
‘To introduce’ and ‘for introducing’	‘Pour/afin de introduire’ and ‘pour l’introduction de’

17 Making the passage from ‘present participle’ to ‘the + noun + of’ in English and from ‘infinitive verb’ to ‘article + noun + de’ in French.

English	*French*
‘Through using this mechanism’ and ‘through the use of this mechanism’	‘Dans le but d’utiliser’ and ‘dans le but de l’utilisation de’

18 Making the passage from ‘in + adjective’ to ‘adverb’ in English and from ‘en + adjective’ to ‘adverb’ in French.

English	*French*
‘In particular’ and ‘particularly’	‘En particulier’ and ‘particulièrement’

19 Making the passage from ‘verb + no + noun’ to ‘does/do + verb + any + noun’ in English and from ‘ne + verb + aucun + noun’ to ‘ne + verb + pas + article + noun’ in French.

English	*French*
‘Which requires no alignment’ and ‘which does not require any alignment’	‘Qui ne requiert aucun alignement’ and ‘qui ne requiert pas un alignement’

20 Making the passage from ‘who/that/which + verb’ to ‘present participle’ in English and from ‘qui + verb’ to ‘present participle’ in French.

English	*French*
‘Who interact’ and ‘interacting’	‘Qui interagit’ and ‘interagissant’

21 Making the passage from ‘verb + of’ to ‘verb + about’ in English and from ‘verb + de’ to ‘verb + autour’ in French.

English	*French*
‘Think of’ and ‘think about’	‘Penser de’ and ‘penser autour’

22 Substituting some words by some ones utilised commonly for avoiding repetitions.

English	*French*
‘From ontological issues to political ones’ and ‘from ontological issues to political issues’	‘De problèmes ontologiques à ceux politiques’ and ‘de problèmes ontologiques aux problèmes politiques’

23 Making the passage from ‘for + noun + to + verb (infinitive form)’ to ‘that + noun + verb conjugated’ in English and from ‘pour + noun + de + verb (infinitive form)’ to ‘que + noun + verb conjugated’ in French.

English	*French*
‘It is essential for the language system to ensure’ and ‘it is essential that the language system ensures’	‘Il est esseniel pour un système linguistique d’assurer’ and ‘il est essentiel qu’un système linguistique assure’

24 Making the passage from ‘better + noun’ to ‘noun + with + better + quality’ in English and from ‘bonne + noun’ to ‘noun + de + bonne + qualité’ in French.

English	*French*
‘It leads to better translation’ and ‘it leads to translation with better quality’	‘Qui donne une bonne traduction’ and ‘qui donne une traduction de bonne qualité’

25 Making the passage from ‘verb1 + to + verb2 + noun’ to ‘verb1 + in + the + action noun of verb2 + of + noun’ in English and from ‘verb1 + à + verb2 (infinitive form) + noun’ to ‘verb1 + à/au + action noun of verb2 + de + noun’ in French.

English	*French*
‘Many actors collaborate to develop an embedded system’ and ‘many actors collaborate in the development of an embedded system’	‘Plusieurs acteurs collaborent à développer un système embarqué’ and ‘plusieurs acteurs collaborent au développement d’un système embarqué’

26 Making the passage from ‘adjective + noun’ to ‘noun + which/that + to be + adjective’ in English and from ‘noun + adjective’ to ‘noun + qui + être + adjective’ in French.

English	*French*
'To highlight informative words' and 'to highlight words which are informative'	'qui identifie les mots informatifs' and 'qui identifie les mots qui sont informatifs'

27 Making the passage from 'verb + able' to 'that/which + can + be + past participle of verb' in English and from 'pouvant être + past participle of verb' to 'the stem of verb + able' in French.

English	*French*
'The data usable for learning' and 'the data that can be used for learning'	'Pouvant être utilisé' and 'utilisable'

28 Updating the position order of some words whose the position does not influence the meaning.

English	*French*
'The nature and the supernatural' and 'the supernatural and the nature'	'La nature et le super-naturel' and 'le super-naturel et la nature'

4.3 Discussion

1 The proposed synonyms extraction solution is simple and practical and may easily replace utilisation of resources such as dictionaries for going to deal with and reutilise available applications such as translation systems.

2 In the case of strong synonyms, the practical aspect of the proposed solution, through using many machine translation systems, is evident so that obtained synonyms have different POS and different number of words.

3 The number of detected synonyms seems to be important (327 to 3,026 for English and 301 to 3,632 for French as depicted in Table 1) and the synonym accuracy (0.8571 in total for English and the average of 0.876 for French as depicted in Table 2) is comparable with the proposed solutions of the literature.

4 The 28 rules given in this work (in Section 4.2) have resulted from the used dataset of ten theses abstracts based on the four considered machine translation systems. Enlarging the dataset for other texts and their associated translated segments, especially with much more translation systems, may generate more rules and yield more weak synonyms.

5 It is easy to see that the returned synonyms are absolute synonyms no matter what its category: weak or strong.

6 Dealing with weak synonyms, based on lexical rules patterns with POS, should receive more attention especially in the machine translation evaluation because the substitution of weak synonyms cannot simply be replaced by stemming or lemmatisation sub-operations.

5 Conclusions

In this paper, we deal with synonyms extraction through introducing a practical method based on adopting some free machine translation systems from the web. We classify synonyms into two kinds: strong synonyms and weak synonyms based on the difference of the considered vocabularies: synonyms with the same vocabularies are qualified as weak synonyms while synonyms with different vocabularies are considered as strong synonyms. From machine translation evaluation point of view, we advise that weak synonyms should receive more attention in synonyms substitution pre-processing operation.

References

Ababneh, A.H., Lu, J. and Xu, Q. (2021) 'New synonyms extraction model based on a novel terms weighting scheme', *Journal of Information and Organizational Sciences*, Vol. 45, No. 1, pp.171–221.

Abbache, A., Meziane, F., Belalem, G. and Belkredim, F.Z. (2018) 'Arabic query expansion using Wordnet and association rules', *International Journal of Intelligent Information Technologies*, Vol. 12, No. 3, pp.51–64.

Al Ameed, H.,K., Al Kebti, S.O., Al Kaabi, A.A., Al Shebli, K.S., Al Shamsi, N.F., Al Nuaimi, N.H. and Al Muhairi, S.S. (2006) 'Arabic search engines improvement: a new approach using search key expansion derived from Arabic synonyms structure', in *IEEE International Conference on Computer Systems and Applications, 2006*, IEEE Computer Society, March, pp.944–951.

AlMaayah, M., Sawalha, M. and Abushariah, M.A. (2016) 'Towards an automatic extraction of synonyms for Quranic Arabic WordNet', *International Journal of Speech Technology*, Vol. 19, No. 2, pp.177–189.

Al-Mathan, R.N. and Al-Khalifa, H.S. (2021) 'Synoextractor: a novel pipeline for Arabic synonym extraction using Word2Vec word embeddings', *Complexity*, Vol. 2021, No. 2, pp.1–13.

Ananthanarayanan, R., Chenthamarakshan, V., Deshpande, P.M. and Krishnapuram, R. (2008) 'Rule based synonyms for entity extraction from noisy text', in *Proceedings of the Second Workshop on Analytics for Noisy Unstructured Text Data*, July, pp.31–38.

Blondel, V.D. and Senellart, P.P. (2002) 'Automatic extraction of synonyms in a dictionary', *Vertex*, Vol. 1, No. 1, pp.1–7.

Bøhn, C. and Nørvåg, K. (2010) 'Extracting named entities and synonyms from Wikipedia', in *2010 24th IEEE International Conference on advanced Information Networking and Applications*, IEEE, April, pp.1300–1307.

Esposito, M., Damiano, E., Minutolo, A., De Pietro, G. and Fujita, H. (2020) 'Hybrid query expansion using lexical resources and word embeddings for sentence retrieval in question answering', *Information Sciences*, Vol. 514, No. 1, pp.88–105.

Inkpen, D. (2007) 'A statistical model for near-synonym choice', *ACM Transactions on Speech and Language Processing (TSLP)*, Vol. 4, No. 1, pp.1–17.

Kazi, L. and Kazi, Z. (2019) 'Using ontology and rule-based reasoning for conceptual data models synonyms detection: a case study', *Journal of Database Management (JDM)*, Vol. 30, No. 1, pp.1–21.

Kumari, M., Jain, A. and Bhatia, A. (2016) 'Synonyms based term weighting scheme: an extension to TF.IDF', *Procedia Computer Science*, Vol. 89, No. 1, pp.555–561.

Lin, D. (1998) 'Automatic retrieval and clustering of similar words', in *36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*, August, Vol. 2, pp.768–774.

Lombardi, M. and Marani, A. (2015) 'SynFinder: a system for domain-based detection of synonyms using WordNet ad the web of data', in *Mexican International Conference on Artificial Intelligence*, Springer, Cham, October, pp.15–28.

Lu, Y., Zhang, C. and Hou, H. (2009) 'Using multiple hybrid strategies to extract Chinese synonyms from Encyclopedia resource', in *2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC)*, IEEE, December, pp.1089–1093.

Miller, G.A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K.J. (1990) 'Introduction to WordNet: an on-line lexical database', *International Journal of Lexicography*, Vol. 3, No. 4, pp.235–244.

Mohammed, N. (2020) 'Extracting word synonyms from text using neural approaches', *Int. Arab J. Inf. Technol.*, Vol. 17, No. 1, pp.45–51.

Muller, P. and Langlais, P. (2011) 'Comparing distributional and mirror translation similarities for extracting synonyms', in *Canadian Conference on Artificial Intelligence*, Springer, Berlin, Heidelberg, May, pp.323–334.

Muller, P., Hathout, N. and Gaume, B. (2006) 'Synonym extraction using a semantic distance on a dictionary', in *Proceedings of TextGraphs: The First Workshop on Graph Based Methods for Natural Language Processing*, June, pp.65–72.

Murata, M., Goda, Y. and Tokuhisa, M. (2014) 'Automatic selection and analysis of synonyms in Japanese sentences using machine learning', in *Proceedings on the International Conference on Artificial Intelligence (ICAI)*, The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), p.1.

Naser-Karajah, E., Arman, N. and Jarrar, M. (2021) 'Current trends and approaches in synonyms extraction: potential adaptation to Arabic', in *2021 International Conference on Information Technology (ICIT)*, IEEE, July, pp.428–434.

Simanosky, A. and Ulanov, A. (2011) 'Mining text patterns for synonyms extraction', in *2011 22nd International Workshop on Database and Expert Systems Applications*, IEEE, August, pp.473–477.

Takeuchi, K. (2008) 'Extraction of verb synonyms using co-clustering approach', in *2008 Second International Symposium on Universal Communication*, IEEE, December, pp.173–178.

Wang, J., Lin, C., Li, M. and Zaniolo, C. (2020) 'Boosting approximate dictionary-based entity extraction with synonyms', *Information Sciences*, Vol. 530, pp.1–21.

Wang, T. and Hirst, G. (2009) 'Extracting synonyms from dictionary definitions', in *Proceedings of the International Conference RANLP-2009*, September, pp.471–477.

Xiang, L., Yu, J., Yu, J., Yang, C., Zeng, D. and Shen, X. (2018) 'A word-embedding-based steganalysis method for linguistic steganography via synonym substitution', *IEEE Access*, Vol. 6, pp.64131–64141.

Yadav, J. and Meena, Y.K. (2016) 'Use of fuzzy logic and wordnet for improving performance of extractive automatic text summarization', in *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, IEEE, September, pp.2071–2077.

Websites

http://translate.google.com (accessed 7 September 2021).

http://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/ (accessed 1 July 2021).

https://www.babylon-software.com/?lang=fr (accessed 7 September 2021).

https://www.bing.com/translator (accessed 7 September 2021).

https://www.online-translator.com/traduction (accessed 7 September 2021).