Color the Word: Leveraging Web Images for Machine Translation of Untranslatable Words

Yana van de Sande(✉) and Martha Larson

Radboud University, Nijmegen, Netherlands
{yana.vandesande,martha.larson}@ru.nl

Abstract. Automatic translation allows people around the globe to communicate with one another. However, state-of-the art machine translation is still unable to capture fine-grained meaning. This paper introduces the idea of using Web image selections in text-to-text translation, specifically for *lacunae*, which are words that do not have a translation in another language. We asked human professional translators to rank Google translate translations of lacunae in German and Dutch. We then compared that ranking with a ranking based on color histograms of Web image data of the words. We found there is viable potential in the idea of using images to address lacunae in the field of machine translation. We publicly release a dataset and our code for others to explore this potential. Finally, we provide an outlook on research directions that would allow this idea to be used in practice.

Keywords: Machine translation · Lacunae · Semantics · Web images

1 Introduction

Translating a text requires transforming the semantics of a source language into a target language. In recent years, machine learning has made dramatic advances in text-to-text translation by exploiting huge amounts of textual training data [2]. Although today's automatic machine translation is highly advanced, the problem of *lacunae* remains unsolved. Lacunae are 'holes in meaning', or words that exist in one language, but have no direct correlate in another. Examples are the Dutch word *gezellig* and the German word *Schadenfreude*, which have no direct translations into English. In this paper, we demonstrate that image data has a potential role to play in the text-to-text translation of lacunae. We introduce the idea that Web image search has potential to extend automatic text-to-text translation in order to address the persisting challenge of lacunae.

Informally, lacunae are considered to be untranslatable words. However, this characterization does not reflect the reality of translation, which cannot simply skip difficult words, but requires the meaning of the source language to be fully and faithfully represented in the target language. When human translators

B. Þór Jónsson et al. (Eds.): MMM 2022, LNCS 13141, pp. 131–138, 2022.
https://doi.org/10.1007/978-3-030-98358-1_11

encounter lacunae in their source language, they are forced to choose the best fitting words in their target language. Their choice should bring the meaning of the source text across to the reader without losing the nuance expressed by original words. For this reason, the difficulty of translating lacunae can be considered to lie in the nature of the words and in the translators' knowledge of the target language, rather than in untranslatability [16].

Our proposition is that image data collected via Web search has the potential to capture knowledge of language that can complement what textual training data contributes to automatic machine translation. Specifically, we argue that images can supply information on the nuances of meaning that are necessary to identify the best fitting words of a target language that should be used to translate a lacuna in the source language. We demonstrate that color-based representations derived from sets of images naïvely collected using a general Web search engine is already surprisingly effective in picking out best-fit targets from among candidate translations.

The idea has social impact due to its potential contribution to intercultural communication. Further exploration of this idea can help improve machine translation, improve our understanding of semantic information in visual data and establish color as a feature of semantics. Improving machine translation can have positive effects for communication between people around the world. Furthermore, we hope to start a conversation on the role of multimedia data, in this research image data, in machine translation research and semantics.

The paper is organized as follows. First, we provide a brief overview of related work demonstrating why our idea can be considered both novel and brave. Next, we describe the set up of our experiment, which compares the ability of image-based features to rank translation candidates with ground truth supplied by a team of expert translators. Finally, we discuss our findings and the implications of our basic experiment for the wider use of image data for the enhancement of text-to-text translation in the future. With this paper we also publicly release a paper repository containing related resources, which will support other researchers in understanding, confirming, and extending our findings.[1] The repository includes our list of lacunae, ground truth (expert-translator rankings of translations), code, and also sets of images that illustrate particular cases.

2 Motivation and Background

2.1 Multimedia Analysis

Multimedia research has recently achieved promising results in the area of multimodal machine translation [17]. In this task, the source text is accompanied by visual content, which is leveraged to improve translation. Our idea is different because it focuses on text-to-text translation in the case of no accompanying visual input and it zeros in on the most challenging cases, lacunae. Intuitively, capturing the nuance of meaning as expressed in images could be expected to

[1] https://github.com/yanavdsande/ColorsOfMeaning.

require sophisticated technology capable of detailed image interpretation. For example, Borth et al. [1] pioneered research on sentiment-related visual ontologies, which directly models language nuance. Because our idea is using image data returned by a Web search engine, we can already access the nuance of image meaning. Detailed understanding of the relationship between image and text that is exploited by the search engine is not necessary.

Image retrieval research often builds on the assumption that images with similar color distributions are semantically similar [7]. Work that demonstrates the relationship between global color features and image content includes [8,13]. In this paper, we query a search engine with lacunae and represent the resulting image set as a color-based vector. Our experimental results suggest that this simple representation is enough to capture a semantic signal that discriminates different possible translations for lacunae, laying the groundwork for investigating more sophisticated image representations in the future.

2.2 Other Research Fields

The idea that image content has a contribution to make to translation is inspired by work in other fields that has indicated that visual features play a role in human translation ability. Neurological studies suggest a close cooperation between brain areas responsible for language processing and visual pathways, especially during translation [9,10]. Studies from the field of communication science show that learning a new language is easier when words are combined with pictures [12]. In addition, psychological research on memory and language proficiency reveal a better performance in remembering words when language is combined with visual stimuli [3,4]. Finally, cognitive models suggest visual information plays a role during sense-giving and language comprehension [7,14]. These studies from different fields suggest that visual information plays a role in awarding meaning to words and could, for these reason, potentially play an important role in text-to-text translation.

3 Experimental Investigation

An experiment was conducted where German and Dutch lacunae were translated by Google Translate. These translations occurred in a ranking which Google Translate considered the best to worst translation option. The translations were shuffled and human professional translators were asked to rank the translations from best to worst to their best knowledge. In addition, a new ranking was created by comparing the color histogram of the source words with color histograms of the target words. Both the Google Translate ranking and the color-based ranking were compared with the ground truth ranking.

3.1 Dataset

A dataset of lacunae with target language (English) was constructed using EUNOIA, a dictionary for lacunae. 22 German and 14 Dutch lacunae were selected. These words were translated through Google translate. Google translate is the most-used and well-known translator machine [6,11,18] using state-of-the-art algorithms, hence seemed the obvious choice. This procedure resulted in a total of 15 lacunae with translations, resulting in 52 translations for 10 German lacunae and 31 translations for 5 Dutch lacunae (N = 113). Words with a one-translation option from Google translate (N = 32) were removed from the first analysis and used as a control for the results through a comparison of the quality score of the translators and the dissimilarity score of the color histograms later. On all words of this selection a Google image search was conducted, collecting 100 images per word, using an image downloader extension.[2] N = 100 was chosen to lower the chance that image sets were lopsided in terms of factors not important for translation (such as contextual or cultural variation). The use of Google image search also allows us to demonstrate that it is not necessary to construct a carefully tailored dataset in order to measure the ability of Web images to support the translation of lacunae. Note that the search results could include both images and graphics. Also, we have no insight in whether Google translate and Google images may share common training data.

To assure search history and cookies did not affect the translations or the image selection, multiple different browsers were used, a private VPN was set up for every search, and the source language was selected prior to translation [5]. In addition, the browser was not logged in to an account to prevent personalization. Waiting 10 min before entering a new search prevented a carry-over effect from the previous search [15]. After image collection, they were resized to contain the same number of pixels per image. Of the collection of these images a color distribution was created by extracting the pixel RGB values and sum them, resulting in one color histogram per word (Fig. 1). More examples can be found on this paper's github.

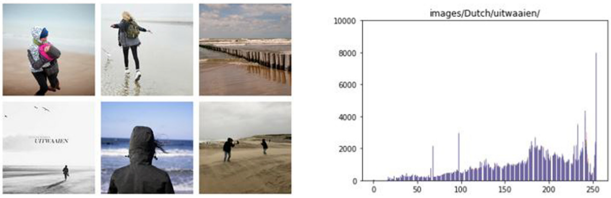

Fig. 1. A selection of images from the word "uitwaaien" and its color histogram

The dissimilarity of the different histograms was calculated using the Chi-Square of the R-channel, G-channel and B-channel between source word and translation, resulting in three measurements of dissimilarity per word-translation combination. To get one value, the three values were averaged. Note that we chose a Chi-Square measure above a KL-divergence since we are in essence not dealing with probability distribution and did not normalize to a range of $[0, 1]$.

[2] https://bulkimagedownloader.com.

$$\sum_{i=1}^{n} \frac{\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{y_i}}{n} \tag{1}$$

In Eq. 1, x_i is the color histogram (with regards to either 'R', 'G' or 'B') of the source-word, y_i is the color histogram of the target-word, n is the number of target words by Google translate.

3.2 Survey and Expert Ground truth

A selection of human professional translators, who are assumed to be expert in the field of human translation, were asked to fill out a survey on translating lacunae. The survey consisted of two parts: part 1 ranking questions in which the translator was asked to rank translation options given by Google translate. Part 2 consisting of a scoring task to judge translation options on their quality. This quality scored will later be used in a second analysis where the score is compared with the chi-square score in regression analysis. The second part functioned as a double control of the first part and as data collection for future research. A Spearman rho test shows a high similarity ($min.Rs(26) = 0.989$, $p < 0.001$, max. $Rs(26) = 1.00$, $p < 0.001$) between the rankings by different translators ($n = 21$). To create a ground truth, the rankings were summed and standardized. The resulted rankings were manually checked by the researcher to see if averaging the rankings did not result in a case of a tied ranking. This was not the case.

3.3 Results

Comparing the ground truth with the rankings of translations by the color histogram, we find that there is a significant relationship between rankings made by human translators and rankings made based on color histograms ($Rs(26) = 0.442$, $p = 0.024$), demonstrating the viability of our idea. When comparing this result to the rankings created by Google Translate, we see that Google translate does outperform the color histogram with a slightly higher Spearman Rho by 0.169, which was significant as well ($p < 0.001$). The second analysis was not significant. Taking a closer look at the data from the first analysis, we do see multiple specific cases where the color histogram translation better approached the human ranked value than did the Google translate algorithm. These specific cases can be found in the paper repository.

4 Discussion and Outlook

The results of our experimental investigation reveal the potential of Web images as a source of information complementary to the textual training material already exploited by Google Translate. We have seen that Google Translate performs already well in ranking candidate translations of lacunae but there is still much to gain. Although the Chi-Square similarity between color histograms shows not to be a good predictor for translation quality, we have seen cases in which our color-based ranking approach has outperformed Google Translate. Our observations support the main claim of our paper, that images on the Web have the potential to extend automatic text-to-text translation. In this section, we provide a discussion and mention future work that can build on our findings.

4.1 Images on the Web

A strength of our approach is that image understanding is not necessary to leverage images to support translation. However, individual cases lead us to believe that introducing additional intelligence into the selection of images could yield benefits. Specifically, when studying the collected images, we noticed that some of the images contained either textual information or depicted an instance that was named after the target-word but was less clearly related to the word itself. This could both be considered an advantage since these images bring connotations and/or related semantics to the equation as a disadvantage since it lacks the representation of a clear meaning.

To overcome this disadvantage, it would be helpful to discover how to preselect images most suitable for supporting translation. We recommend to use a pre-checked database or to use coders to select images that are related to the words. Using a pre-checked database would also solve to problem of images that had a focus on other semantic properties. For example, an image of a butterfly in the sky. In this example the blue sky causes blue to be over-represented in the color histogram, overshadowing the colors of the butterfly that is actually the subject of the image. Example images can be found on this paper's github.

4.2 Towards the Future of Machine Translation

Regarding the future of machine translation, one can take inspiration from our idea looking at it from different fields. We believe it would be interesting to replicate this study by integrating context through word combinations; an example of this can be found in the translation of "cosy cottage" that should be translated in Dutch as "knus huisje" instead of "gezellig huisje" which is the translation Google translates raises. In addition, we suggest future work takes into account the possible relationship between the data used to train the machine translation and image search engine used, in order to understand the specific contribution of visual content in isolation.

Another direction could be taken from an engineering point of view. Our research shows the potential of making use of visual information during translation. This can, for instance, be used for the creation of an extension that includes a depiction of the target word and its translations so the user can check whether the translation covers what they try to communicate.

Concluding, exploring the potential of visual information resulted in the creation of an open access database consisting of lacunae and their translations by human translators so others could explore the potential of this idea. It showed results that indicate color could be used to enhance the translation of lacunae, and it provided research directions for the future.

References

1. Borth, D., Ji, R., Chen, T., Breuel, T., Chang, S.F.: Large-scale visual sentiment ontology and detectors using adjective noun pairs. In: Proceedings of the 21st ACM International Conference on Multimedia, MM 2013, pp. 223–232 (2013)
2. Brants, T., Popat, A.C., Xu, P., Och, F.J., Dean, J.: Large language models in machine translation. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), pp. 858–867 (2007)
3. Carpenter, S.K., Geller, J.: Is a picture really worth a thousand words? Evaluating contributions of fluency and analytic processing in metacognitive judgements for pictures in foreign language vocabulary learning. Q. J. Exp. Psychol. **73**(2), 211–224 (2020)
4. Carpenter, S.K., Olson, K.M.: Are pictures good for learning new vocabulary in a foreign language? Only if you think they are not. J. Exp. Psychol. Learn. Mem. Cogn. **38**(1), 92 (2012)
5. Hannak, A., et al.: Measuring personalization of web search. In: Proceedings of the 22nd International Conference on World Wide Web, pp. 527–538 (2013)
6. Kreisha, A.M.: 3 contenders for the title of "best free online translator". https://www.fluentu.com/blog/best-free-online-translator. Accessed 10 Oct 2021
7. Lavrenko, V., Manmatha, R., Jeon, J., et al.: A model for learning the semantics of pictures. In: Advances in Neural Information Processing Systems 16 (2003)
8. Lux, M., Riegler, M., Halvorsen, P., Pogorelov, K., Anagnostopoulos, N.: LIRE: open source visual information retrieval. In: Proceedings of the 7th International Conference on Multimedia Systems, MMSys 2016 (2016)
9. Mayer, K.M., Yildiz, I.B., Macedonia, M., von Kriegstein, K.: Visual and motor cortices differentially support the translation of foreign language words. Curr. Biol. **25**(4), 530–535 (2015)
10. Oldfield, R.C.: The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia **9**(1), 97–113 (1971)
11. Otachi, E.: 12 best online translators to translate any language. https://helpdeskgeek.com/free-tools-review/12-best-online-translators-to-translate-any-language Accessed 10 Oct 2021
12. Plass, J.L., Jones, L.: Multimedia learning in second language acquisition. In: The Cambridge Handbook of Multimedia Learning, pp. 467–488 (2005)
13. Riegler, M., Larson, M., Lux, M., Kofler, C.: How 'how' reflects what's what: Content-based exploitation of how users frame social images. In: Proceedings of the 22nd ACM international conference on Multimedia, MM 2014 (2014)
14. Roy, D.K., Pentland, A.P.: Learning words from sights and sounds: a computational model. Cogn. Sci. **26**(1), 113–146 (2002)
15. Salehi, S., Du, J.T., Ashman, H.: Examining personalization in academic web search. In: Proceedings of the 26th ACM Conference on Hypertext & Social Media, pp. 103–111 (2015)
16. Sankaravelayuthan, R.: Lexical gaps and untranslatability in translation. Lang. India **20**(5), 56 (2020)

17. Song, Y., Chen, S., Jin, Q., Luo, W., Xie, J., Huang, F.: Enhancing neural machine translation with dual-side multimodal awareness. IEEE Trans. Multimedia (2021)
18. Writtenhouse, S.: The 10 best online translators you can use in the real world. https://www.makeuseof.com/tag/best-online-translators Accessed 10 Oct 2021