IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol. 003, LIIR056
DOI: https://doi.org/10.52731/liir.v003.056

CLE-UMA – A Creative Learning Environment Using Multimodal Analogies

Werner Winiwarter * , Bartholomäus Wloka †

Abstract

In this paper, we present a novel approach for applying multimodal analogies to creative learning scenarios. We have chosen Japanese language learning, in particular kanji acquisition, as the first application domain for our research. Our solution is based on a kanji dictionary which we enriched with visual, compositional, and semantic information. We have manually assigned images to 3,500 kanji and elaborately annotated the mappings from the lexical data to its visual representation. This knowledge enables us to find high-quality analogies for associations between the visual and textual dimensions. The analogies are integrated into our Web-based contextual language learning environment to empower several creative learning applications through augmented browsing technology including exercises, quizzes, and educational games.

Keywords: computer-assisted language learning, creative learning, lexical semantics, multimodal analogies, natural language annotation.

1 Introduction

Learning a foreign language as an adult can be laborious, often tedious, and sometimes frustrating since results tend to come so slowly that it feels like we are not making any progress at all [1][2]. This can be magnified by the fact that we tend to fixate on obtaining results. Traditional ways of instruction often dictate that there is no other way than to go through a painful process of memorization of vocabulary and rules, *before* we know their use. It is almost like memorizing a random set of characters or numbers, while having to trust the teacher that this will make sense later. This is particularly the case for English-Japanese [3], mainly due to the drastically different writing system, which needs to be learned as soon as possible to make progress [4]. We present a method to ease this process of deciphering and necessary memorization with a novel approach which includes multimodal analogies in the broader context of *creative learning* [5].

Creative learning gives the learners a tool box, rather than telling them how to use individual tools [6]. They can use their own creativity to shape their own learning path.

* University of Vienna, Faculty of Computer Science, Vienna, Austria
† University of Vienna, Department for Translation Studies, Vienna, Austria

Furthermore, we let the students choose their own learning context, so they can work on a topic that is interesting to them. Though it might sound trivial, this is a key component to keep up motivation over an extended period of time. Giving the students material of their choice combined with exciting ways of instruction and exploration of that content keeps a high level of engagement and enjoyment [7].

The five main reasons why creative learning is so important can be summarized according to [8] as follows:

1. it **stimulates problem solving**, which can be further enhanced by various *gamification* techniques;
2. the inclusion of the learners in the creation of the learning path puts them in control over their progress and approach, which **develops critical thinking** and gives an extra sense of achievement that is vital for continued motivation;
3. the ability to make mistakes when trying out new things is vital in the learning process since it **promotes risk-taking** and fosters the understanding that mistakes ultimately lead to success; this helps to remove the "barrier" that makes many beginners remain silent due to the fear of "saying it wrong";
4. it **builds a curious mindset**; curiosity about the subject matter and the method of learning encourages the students to try new things and discover patterns, rules, associations, analogies, etc.;
5. it **increases confidence levels** of students so that they are more likely to apply this content in everyday life, which in turn boosts motivation to create and explore more exciting learning content.

We address all of these aspects with the development of our creative learning environment **CLE-UMA** with the aim to contribute a vital stimulus for second language acquisition of Japanese. Our target audience are students of Japanese at almost any level of proficiency. One of the reasons that makes Japanese language so difficult to learn is the writing system. It uses logographic characters in addition to two syllabaries. The logographic characters *kanji* pose a significant challenge [9]. Each of these characters usually has several possible meanings and pronunciations depending on the usage context. What makes the study of those characters even more daunting is their sheer number. A standard kanji dictionary includes more than 10,000 of these characters and over 50,000 possible words built from them. These characters are taught incrementally to Japanese native speakers from early childhood throughout their high school education (80-200 kanji per school year) and additional characters are learned by Japanese native speakers throughout their entire adult live. Obviously, it is a formidable challenge to shorten this steady and incremental learning process to a few lessons for foreign language students who have to study both the characters and the language at the same time. It is important to mention that whereas it is possible to avoid kanji by representing text with only the two syllabaries, the study of the Japanese language cannot progress past a certain very basic level without the understanding and use of kanji. This makes kanji vital for a proper and deep understanding of Japanese which motivated us to address the challenges of kanji study with our novel approach using *multimodal analogies*.

Analogies are a very intuitive concept and can be applied to teaching and learning [10], machine translation [11], and natural language processing (NLP), e.g. the prediction of unknown morphological variations of words, as shown in [12]. Analogies also can be used very effectively across languages. Similarities (or ratios) between objects or concepts are universal, hence they can facilitate a language-independent bridge between them. For the sake of simplicity we present this concept in the most intuitive form of the formal notation of two ratios, such as:

chair : armchair :: stool : ***X***.

Even if we do not know the value of X, we start to form ideas and concepts immediately upon extracting the similarities and differences of the properties of “chair” and “armchair”. Once we substitute X with “curule seat”, we quickly get an idea what a curule seat is, even if we did not know this term before. This is due to the ratios and similarities of the other three obvious terms in the analogy structure:

chair : armchair :: stool : **curule seat**.

The use of these intuitive relations helps to construct and build the new term from already known concepts, rather than merely learning a new vocabulary equivalent in two languages. The process of reconstruction and association greatly improves the long term memorization effect. We exploit such associations between the visual and textual dimension with our multimodal analogies as a novel key component of our creative learning environment CLE-UMA.

2 Building an Enhanced Kanji Dictionary

As one of the main components of our contextual Japanese language learning environment, we have created an enhanced kanji dictionary, which includes images, structural decompositions, and a detailed cross-media annotation of the mappings from lexical data to visual representations in the depictions. The annotations are semantically grounded by disambiguating word senses through references to concepts in a lexical-semantic network. They form the foundation for our approach to create meaningful multimodal analogies for creative kanji learning.

Figure 1 gives an overview of the linguistic resources that we used to develop our kanji dictionary containing 3,500 fully annotated kanji. We extract all the lexical data from *KANJIDIC2* [13], a freely available lexical resource with detailed information about 13,108 kanji. The images we display on the kanji cards were all collected manually from *Wikipedia* pages. We exclusively use Wikipedia as source not only for licensing reasons but mainly because of the valuable contextual and ontological data. The *IDS data* for the ideographic description sequence is derived from the CHISE project (`www.chise.org`) and freely available at `github.com/cjkvi/cjkvi-ids`. We used *Wiktionary* as supplementary resource to correct more than 1,000 KANJIDIC2 and IDS data entries. Finally, the main semantic reference for our annotations is *WordNet* [14]. We have fully implemented our language learning environment in *SWI-Prolog* [15] because it is the optimal choice for efficient natural language processing solutions as well as the management and retrieval of linguistic data. It also offers a scalable Web server solution to generate dynamic Web content for our augmented browsing interfaces and libraries for dealing with huge XML and RDF files.

The number 166 in Figure 1 is the *radical* number, which is used to look up the kanji in a paper kanji dictionary. There exist 214 kanji that are used as radicals. If the radical appears in the IDS, it is highlighted in red. One essential feature for language students is that they can recursively explore the decomposition of a kanji into its components down to individual radicals or some additional strokes as can be seen in Figure 2. Additional radical kanji are highlighted in orange in the IDS, radical kanji cards are displayed in red,

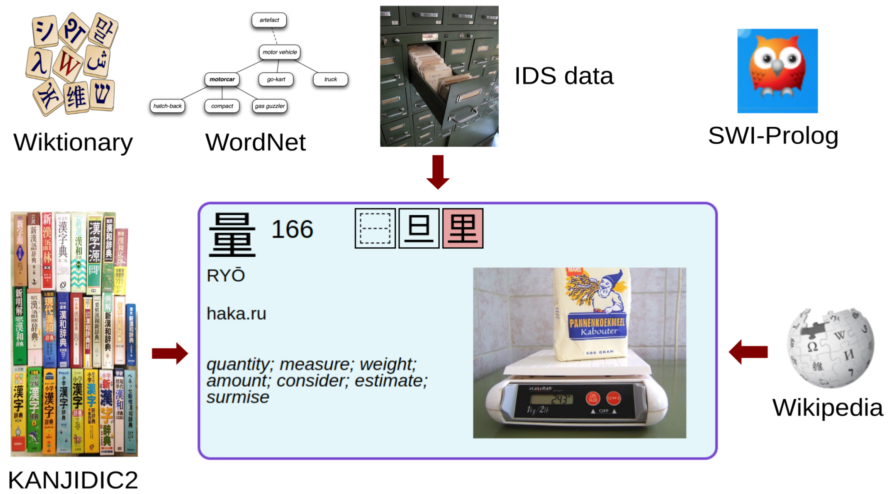


Figure 1: Overview of linguistic resources

and possible variant forms are shown for some kanji. This kind of assembly diagram is very useful for memorizing kanji through creative storytelling, e.g. "if you *plan* to *fathom* the depth of the river, you need a long *rule*" or "you should *follow* the *rule* to use the right *knife* to shuck *shellfish*".

During the time-consuming process of selecting suitable images for kanji cards, we soon noticed that there were a few promising strategies to successfully choose visual representations of abstract concepts. We also realized that in many cases we covered several aspects of the lexical data in one image. Quite often this even happened subconsciously so that it needed another person to become aware of these hidden associations. All these observations led us to think about annotating this mapping process to preserve a formalized representation of the image acquisition task. We discovered that this could be very valuable data for several more advanced applications, especially because we also annotated the particular word sense for each mapping by assigning it to the correct synset in WordNet.

Figure 3 shows an example of the mappings for a kanji card. We have converted the verbal descriptions into a more formal annotation format as a set of septuples:

$$\{(\texttt{LexicalData},\texttt{SourceType},\texttt{InformationType},\\ \texttt{TargetRegion},\texttt{DepictionType},\texttt{MappingType},\texttt{Synset})\}$$

For instance, the first mapping in Figure 3 would be translated into:

$$(\text{grope},\text{gloss},\text{activity},\text{main motif},\\ \text{graphic artwork},\text{execution},\text{grope.v.02})$$

We actually use one-letter codes for the values of the five features to create annotations for all 3,500 kanji in our dictionary resulting in a total number of 4,282 annotated mappings.

For the efficient generation of our multimodal analogies, we performed the following simplifications and aggregations on the original annotation (see [16] for more details):

1. we only use the first primary mapping for each kanji because it carries the main semantic contribution;

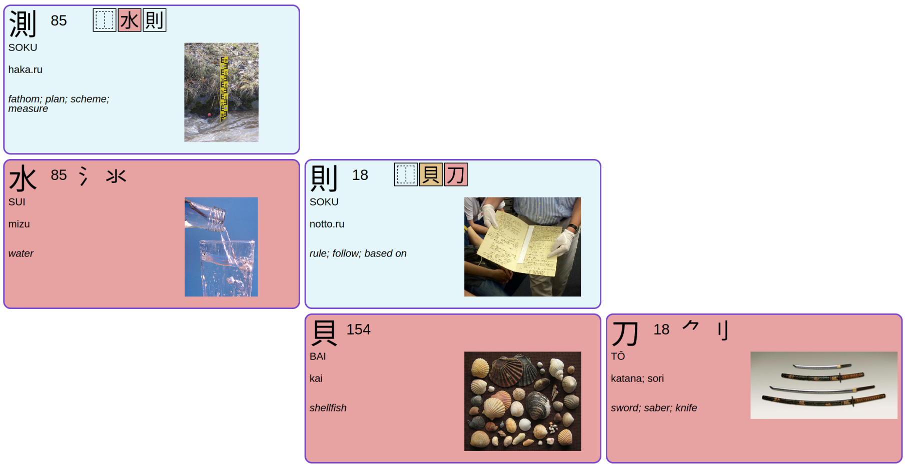


Figure 2: Example of decomposition of a kanji

2. accordingly, since the first mapping only concerns glosses, we can omit the *source type* from the annotation;
3. we also remove the information about the *target region* because it is mainly of importance to the additional mappings for radicals, components, etc.;
4. we merge several fine-grained feature values for which the distinction is of rather low relevance for our purposes.

Therefore, we ended up with the following optimized annotation:

$$(\texttt{Gloss}, \texttt{InformationType}, \texttt{DepictionType}, \texttt{MappingType}, \texttt{Synset})$$

For example, the first mapping in Figure 3 “mapping gloss of type activity (A) to main motif of the graphic artwork (G) depicting its execution (E)” is represented as:

$$(\text{grope}, A, G, E, \text{grope.v.02})$$

We use 22 different feature values, which are introduced in the next section and summarized later in Table 1.

3 Multimodal Analogies

In this section we present several examples of different types of analogies to illustrate their utilization. We divide the kanji cards in four *analogy sets* according to the *information type*, which are discussed in the following separate subsections. Within an analogy set we use the *depiction type* and *mapping type* to further filter down the kanji cards to retrieve similar pairs. Finally, we use the WordNet concept hierarchy, in particular the *path distance similarity* and the *lowest common hypernyms* for the two synsets, to obtain kanji cards that are sufficiently similar but at the same time not too close in meaning or even identical or synonymous.

3.1 Analogies for Tangible Objects

Figure 4 shows an analogy for a *tangible object* (T) using *real depictions* (R) depicting the objects *directly* (D).

Figure 3: Example of mappings

(strawberry,T,R,D,strawberry.n.01) (apricot,T,R,D,apricot.n.02)

Figure 4: Example of analogy for feature values (T,R,D)

(bleach,T,R,T,bleaching agent.n.01) (sparkling water,T,R,T,soda_water.n.01)

Figure 5: Example of analogy for feature values (T,R,T)

The path distance similarity for the two synsets is 0.25. The lowest common hypernym for the two synsets is `edible_fruit.n.01`. In Figure 4 and all the following examples we display the two images and the gloss information. If we want to use the analogies for language learning, we would replace the display of the gloss "apricot" with an input field to ask the student for the correct solution. Acceptable answers could be extended to synonyms by using the lemma names for the WordNet synset. After this quite straightforward, simple first example, Figure 5 shows a more elaborate case. Here, the identification of the visual representation is assisted by *textual* (T) information, i.e. the brand name.

3.2 Analogies for Properties

Properties (P) are much more difficult to map to visual representations than tangible objects. The most common strategy is to find objects with the same *properties* (P). For instance, Figure 6 shows an example using colors.

A second very promising avenue is to use *zoological subjets* (Z), i.e. associations to animal behavior. Figure 7 uses the preconception that deer are rather cautious whereas donkeys are stubborn.

A similar approach is to use *well-known personalities* (W), e.g. scientists, actresses, politicians, sports persons, authors, etc. Figure 8 shows an example attributing Nelson Mandela with wisdom and Lady Diana with kindness. Finally, Figure 9 illustrates the use of personality traits of characters from popular *narrations* (N) depicted in *graphic artworks* (G). As can be seen, Cinderella leads a wretched life, whereas the Cowardly Lion in the children's novel "The Wonderful Wizard of Oz" is weak until he receives a medal for his bravery and conquers his self-doubt.

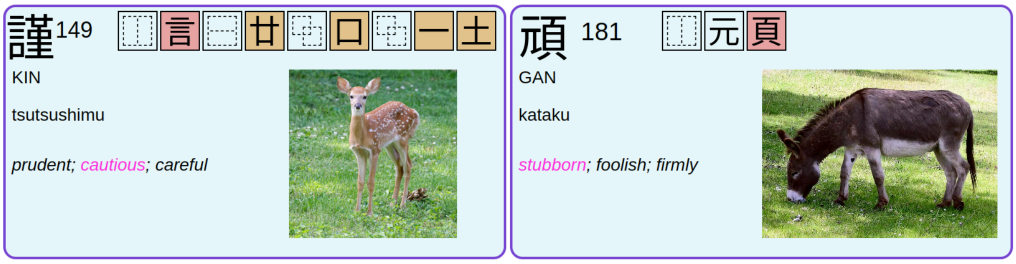
3.3 Analogies for Activities

The most common visualization of *activities* (A) is to show a scene or situation at the time when the activity is *executed* (E). In Figure 10 we use two *film scenes* (F) for this purpose. In the first one from "Meat Market 3", the undead draw near in a threatening manner; in the second one from "Rebecca of Sunnybrook Farm", we can see how a student is tormented by her classmate.

The second way to represent activities is to display their *result* (R), i.e. the state after completing the activity successfully. In Figure 11, we can witness a nicely wrapped present and a neatly mowed lawn.

(white,P,R,P,white.a.01) (red,P,R,P,red.s.01)

Figure 6: Example of analogy for feature values (P,R,P)



(cautious,P,R,Z,timid.n.01) (stubborn,P,R,Z,stubborn.a.01)

Figure 7: Example of analogy for feature values (P,R,Z)

A very convenient way to convey the intended meaning of an image is to use the *literal* (L) mention of the concept as textual string. Figure 12 shows two *written materials* (W), a Canadian World War I recruiting poster and a cover page for a fantasy magazine.

3.4 Analogies for Intangible Objects

The most complex kind of concepts to visualize are *intangible objects* (I). In general, *associations* (A) between the image and the concept have to be identified. Graphic art is one of the preferred media through which artists have tried to express abstract ideas since the dawn of human civilization. Figure 13 gives an example of such an analogy.

Another long tradition of transferring immaterial objects to perceivable artefacts are

(wise,P,R,W,wise.a.01) (kind,P,R,W,kind.a.01)

Figure 8: Example of analogy for feature values (P,R,W)

惨 61 心 参
SAN; ZAN
miji.me; ita.mu; mugo.i
wretched; disaster; cruelty; harsh

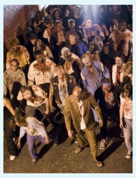
(wretched,P,G,N,miserable.s.05)

弱 57 弓 冫 弓 冫
JAKU
yowa.i; yowa.ru; yowa.maru; yowa.meru
weak; frail

(weak,P,G,N,weak.s.11)

Figure 9: Example of analogy for feature values (P,G,N)

寄 40 宀 奇
KI
yo.ru; -yo.ri; yo.seru
draw near; stop in; bring near; gather; collect; send; forward



(draw near,A,F,E,approach.v.01)

苛 140 艸 可
KA
iji.meru; saina.mu; irada.tsu; karai; komakai
torment; scold; chastise

(torment,A,F,E,torment.v.02)

Figure 10: Example of analogy for feature values (A,F,E)

包 20 勹 己
HŌ
tsutsu.mu; kuru.mu
wrap; pack up; cover; conceal

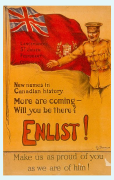
(wrap,A,R,R,envelop.v.01)

乂 4 丿 丿
GAI; KAI; GE
osa.meru; ka.ru
mow; cut grass; subdue

(mow,A,R,R,mow.v.01)

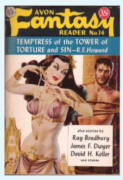
Figure 11: Example of analogy for feature values (A,R,R)

募 19 莫 力
BO
tsuno.ru
recruit; campaign; gather (contributions); enlist; grow violent



(enlist,A,W,L,enlist.v.01)

唆 30 口 夋
SA
soso.ru; sosonoka.su
tempt; seduce; instigate; promote



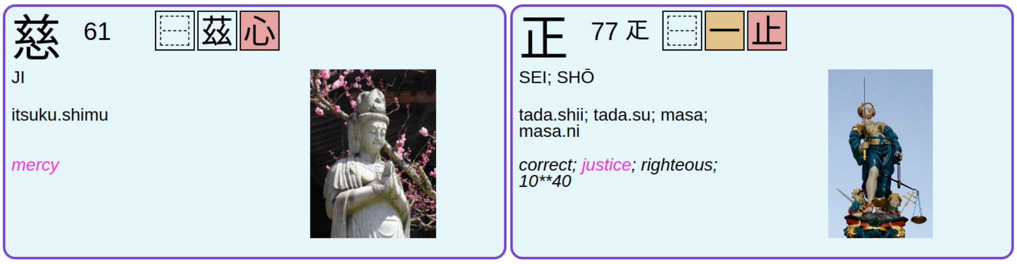
(tempt,A,W,L,tempt.v.03)

Figure 12: Example of analogy for feature values (A,W,L)

(melancholy,I,G,A,melancholy.n.01) (dream,I,G,A,dream.n.01)

Figure 13: Example of analogy for feature values (I, G, A)

monumental sculpture (M) and any other form of three-dimensional visual arts. Anthropomorphic *metaphors* (M), in particular personifications, were already used by ancient advanced civilizations for this purpose. In many polytheistic religions, abstract concepts were also attributed to deities. Figure 14 shows two examples: the left image depicts Guanyin bodhisattva (in Japanese Kannon) who is also named “goddess of mercy”, the right image goddess Justitia with the symbolic items balance, sword, and blindfold.



(mercy,I,M,M,mercifulness.n.02) (justice,I,M,M,justice.n.01)

Figure 14: Example of analogy for feature values (I, M, M)

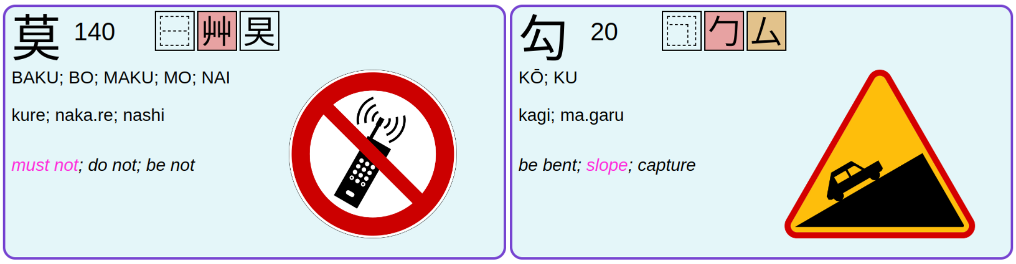
The final example uses *icons* (I), also called pictograms. Their meaning is interpreted through their resemblance to a physical object. Both images in Figure 15 represent *signs* (S), the left one is a prohibitory sign, the right one a warning sign used as road sign in Poland.

For easier reference, we add an overview of all feature values introduced in this section in Table 1.

4 Conclusion

In this paper, we have presented CLE-UMA, a creative learning environment using multimodal analogies. As backbone of the system we have developed an enhanced kanji dictionary with detailed visual, compositional, and semantic annotations. We have described our motivation and the steps involved in building this lexical resource. Our main contribution is the retrieval of high quality analogies based on a formal representation of semantically grounded cross-media mappings from the textual to the visual dimension.

Our future research will focus on editorial work and a comprehensive annotation guide



(must not,I,S,I,forbid.v.01) (slope,I,S,I,gradient.n.02)

Figure 15: Example of analogy for feature values (I,S,I)

Information Type	**Depiction Type**	**Mapping Type**
A activity	*F* film scene	*A* association
I intangible object	*G* graphic artwork	*D* direct
P property	*M* monumental sculpture	*E* execution
T tangible object	*R* real depiction	*I* icon
	S sign	*L* literal
	W written material	*M* metaphor
		N narration
		P property
		R result
		T textual
		W well-known personality
		Z zoological subject

Table 1: Overview of feature values

before publishing our resource. We will also address the challenge of semantic disambiguation at both the textual and visual dimension, e.g. by providing mappings to several images or using image segmentation techniques to highlight regions of interest.

We already envision many interesting use cases. For instance, we have developed two educational strategic card games for which we replace the original thematic cards with suitable kanji cards chosen from a specific kanji collection, e.g. the kanji contained in the currently visited Web page. Previously, we performed the mapping from kanji cards to game cards based on radicals, a coarse solution with some shortcomings that proved unsatisfactory. With our multimodal analogies we have now a much better foundation to choose kanji cards based on feature value agreements and synset similarity.

All implementations will be evaluated in graduate level courses with students from translation studies. We will follow recent suggestions in evaluation approaches, such as [17], by putting the learners in the center of the evaluation process and giving them an active role in the further development of our contextual language learning environment. We will collect feedback with surveys at the beginning, during, and at the end of the evaluation to include critique and suggestions in the continued dynamic development of CLE-UMA.

References

[1] C.J. Doughty and M.H. Long, The Handbook of Second Language Acquisition, Blackwell Publishing Ltd, 2003.

[2] B. VanPatten, M. Smith, and A.G. Benati, Key Questions in Second Language Acquisition, Cambridge University Press, 2020.

[3] H. Matsumoto, "Peak learning experiences and language learning: A study of American learners of Japanese," Language Culture and Curriculum – LANG CULT CURRIC, vol. 20, 2007, pp. 195–208.

[4] S. Paxton, "Kanji matters in a multilingual Japan," The Journal of Rikkyo University Language Center, vol. 42, 2019, pp. 29–41.

[5] L.S. Watts and P. Blessinger, Creative Learning in Higher Education, Routledge, 2017.

[6] D. Davies et al., "Creative learning environments in education – A systematic literature review," Thinking Skills and Creativity, vol. 8, 2013, pp. 80–91.

[7] M. Kubota, "Post study abroad investigation of kanji knowledge in Japanese as a second language learners," System, vol. 69, 2017, pp. 143-152.

[8] ELM Learning, The Difference Between Creativity vs. Creative Learning (and How to Harness Both), elmlearning.com/blog/creativity-vs-creative-learning/, Accessed: 2022-08-18.

[9] Y. Mori, "Review of recent research on kanji processing, learning, and instruction," Japanese Language and Literature, vol. 48, no. 2, 2014, pp. 403–430.

[10] P. Aubusson, D. Treagust, and A. Harrison, Learning and Teaching Science with Analogies and Metaphors, The World of Science Education: Handbook of Research in Australasia, Sense Publishers, 2009, pp. 199–226.

[11] H. Somers, S. Dandapat, and S.K. Naskar, "A review of EBMT using proportional analogies," EBMT 2009 – 3rd Workshop on Example-Based Machine Translation, 2009.

[12] R. Fam and Y. Lepage, "Morphological predictability of unseen words using computational analogy," CEUR Workshop Proceedings, vol. 1815, 2016, pp. 51–60.

[13] J. Breen, "Multiple indexing in an electronic kanji dictionary," Proceedings of the Workshop on Enhancing and Using Electronic Dictionaries, 2004, pp. 1–7.

[14] G.A. Miller, "WordNet: A lexical database for English," Commun. ACM, vol. 38, no. 11, 1995, pp. 39–41.

[15] J. Wielemaker, T. Schrijvers, M. Triska, and T. Lager, "SWI-Prolog," Theory and Practice of Logic Programming, Cambridge University Press, vol. 12, no. 1-2, 2012, pp. 67–96.

[16] W. Winiwarter and B. Wloka, "VISCOSE – A kanji dictionary enriched with VISual, COmpositional, and SEmantic information," Proceedings of the AACL-IJCNLP 7th Workshop on Cognitive Aspects of the Lexicon (CogALex-VII), 2022, pp. 68–77.

[17] H. Heuer and D. Buschek, "Methods for the design and evaluation of HCI+NLP systems," arXiv, vol. 2102.13461 [cs.CL], 2021.