VOLARE – Visual Ontological LAnguage REpresentation

Werner Winiwarter
University of Vienna, Faculty of Computer Science
Währingerstrasse 29, 1090 Vienna, Austria
werner.winiwarter@univie.ac.at

Abstract

In this paper, we introduce a novel meaning representation, which is based on AMR but extends it towards a visual ontological representation. We visualize concepts by representative images, and roles by emojis. All concepts are identified either by PropBank rolesets, Wikipedia page titles, WordNet synsets, or Wikidata lexeme senses. We have developed a Web-based annotation environment enabled by augmented browsing and interactive diagramming. As first application, we have implemented a multilingual annotation solution by using English as anchor language and comparing it with French and Japanese language versions. Therefore, we have extended our representation by a translation deviation annotation to document the differences between the language versions. The intended user groups are, besides professional translators and interpreters, students of translation, language, and literary studies. We describe a first use case in which we use novels by French authors and compare them with their English and Japanese translations. The main motivation for choosing Japanese is the soaring popularity of Japanese courses at our university and the particular challenges involved with trying to master this language.

Keywords: meaning representation, AMR, visual annotation, Web-based annotation environment, multilingual annotation, translation annotation, Japanese

1. Introduction

In recent years, there have been many significant developments in the field of designing meaning representations. The most influential approach has been Abstract Meaning Representation (AMR), which again has inspired a wealth of research work to develop parsers and other tools for AMR.

Building on these great efforts, we have extended AMR towards a multilingual representation by adding a translation deviation annotation. Originally, its main intended use has been within the scope of a more far-reaching international research initiative with the aim to assist interpreters and translators with the task of familiarizing themselves with new domain-specific topics (Wloka et al., 2022).

Beyond that we also target educational applications for students of translation, literary, and language studies. In this context we intend to enable classroom scenarios with individual annotation tasks, where the personal knowledge bases can be compared and aggregated for instructional use.

We have implemented a use case of a Web-based annotation environment, which makes it possible to study novels by French authors and compare them with their English and Japanese translations. Japanese was chosen mainly because of the global manga craze, which led to an unprecedented increase in demand for Japanese language courses and, consequently, technological support.

As an important prerequisite for such a scenario we use English as anchor language and map all concepts to disambiguated unique sense identifiers. Each concept is visually represented by an image. For Wikipedia pages, we allocate and download the image through the corresponding Wikidata entry. Other concepts are associated with images from a collection, which we created in our previous research (Winiwarter and Wloka, 2022). This image database contains currently over 3,500 images from Wikimedia Commons of which more than 60% represent abstract concepts. All images are manually selected and annotated with semantic tags and links to WordNet synsets. We also visually represent all roles by using emojis. For that purpose, we map core roles to suitable thematic roles.

For the rendering at the Web client, we use an interactive diagramming library so that the user can freely edit any aspect of the annotation to offer optimal customizability. For example, the user can update the mapping rules from AMR concepts to uniquely identifiable sense definitions, the links between concepts and links, as well as the links between roles and emojis.

After the successful evaluation of our use case implementation in university courses and with professional translators and interpreters, we will make our Web-based annotation environment freely available at GitLab.

This paper is organized as follows. In Sect. 2 we provide related work on topics relevant for this research including some background on Japanese as far as it is helpful for a better understanding; in Sect. 3 we first discuss the design of the meaning representation and user interface; in Sect. 4 we then describe implementation details; and in Sect. 5 we finish with an outlook towards future work.

5th Workshop on Designing Meaning Representations (DMR 2024) @LREC-COLING-2024, pages 54–65
21 May, 2024.

2. Related Work

2.1. Meaning Representations

The annotation of sentences with *meaning representations* has established itself in the last decade as a thriving research field in computational linguistics (see Abend and Rappoport, 2017). The most influential and most actively promoted approach has been the *Abstract Meaning Representation*[1] (AMR) (Banarescu et al., 2013). There are many parsers available, the best[2] parser being at the moment Lee et al. (2022). The *SPRING* parser (Bevilacqua et al., 2021) can be tried via a Web interface[3], which also offers a nice visualization. One point of criticism concerning AMR's reliance on numbered, not directly interpretable core arguments, is addressed by the *WISeR* meaning representation (Feng et al., 2023), which maps them to thematic roles. There also exist several AMR annotation tools, one recent Web-based solution is *CAMRA* (Cai et al., 2023).

AMR has been recently extended to the *Uniform Meaning Representation*[4] (UMR) (Gysel et al., 2021). It enhances AMR by adding support for other languages (in particular low-resource languages), and a document-level representation capturing intersentential coreference and temporal/modal dependencies. There is an upcoming workshop to kick-start the development of UMR parsers[5].

According to the *UMR guidelines*[6], UMR fully embraces *radical construction grammar* as a theoretical foundation (Croft, 2001, 2022), which was designed with *typological* (Croft, 2002) applicability as main motivation, i.e. to study and classify languages according to their structural features to allow their comparison. Radical construction grammar considers word classes and other syntactic structures as language-specific and construction-specific (Croft, 2023).

2.2. Multimodality

Multimodal enhancements of lexical resources have a long history but only recently gained new momentum due to the strong interest in research on *visual question answering* (VQA) (Lerner et al., 2024) or *multimodal large language models* (MLLMs) (Bewersdorff et al., 2024). One example of an attempt towards a multimodal semantic representation is *VoxML* (Pustejovsky et al., 2016).

Regarding the mapping of images to WordNet synsets, there exists the *ImageNet* collection, which maps ca. 1,000 images to each synset (Deng et al., 2009). Another effort to assign cliparts to synsets was discontinued after illustrating only 581 synsets (Bond et al., 2009). A much more influential resource is *Wikipedia*, which has been increasingly enhanced with visual representations. However, the number of images varies widely across language versions. The most comprehensive recent effort is certainly *BabelNet*[7] (Navigli et al., 2021) with the annotation tool *Babelfy*[8] (Moro et al., 2014) and the latest *BabelPic*[9] (Calabrese et al., 2020) dataset targeting non-concrete concepts.

There also exists a subfield of *cognitive linguistics* dealing with identifying and analyzing language-image relations in multimodal texts, e.g. research on intersemiotic convergence (Hart and Queralto, 2021). One central term in this context is *grounding*, which is interpreted in quite different ways by natural language processing and cognitive science researchers (see Chandu et al., 2021). Whereas natural language processing emphasizes the linking of text to other modalities, cognitive science focuses on how speakers build the common ground to share mutual information. During this cognitive process, a set of abstract symbols acquire meaning through perceptions and situated actions (Chen et al., 2023) in analogy to the concept of *construal* in cognitive linguistics (Langacker, 2008), which accounts for choosing alternative linguistic expressions for expressing the same situation (Divjak et al., 2020).

The use of pictorial illustrations has a long history in language teaching didactics and there exist numerous empirical studies that show their effectiveness at all levels of proficiency, e.g. Tahiri (2020). Nonetheless, to the best of our knowledge, we are not aware of any related work with the aim of creating visual representations of meaning representations of sentences.

2.3. Translation Deviations

While there is an ample supply of tools for translators (Rothwell et al., 2023), there have been comparatively few research efforts on annotating translation deviations. One example is Deng and Xue (2017), who analyzed deviations between Chinese and English texts produced by machine translation. There exists a related research work on creating corpora of machine translated documents with annotated translation errors (Fishel et al., 2012),

[1] https://amr.isi.edu/
[2] https://paperswithcode.com/task/amr-parsing/latest
[3] http://nlp.uniroma1.it/spring/
[4] https://umr4nlp.github.io/web/
[5] https://umr4nlp.github.io/web/UMRParsingWorkshop.html
[6] https://github.com/umr4nlp/umr-guidelines/
[7] https://babelnet.org/
[8] http://babelfy.org/
[9] https://sapienzanlp.github.io/babelpic/

and a more recent work on creating an English-French-Chinese corpus annotated with translation relations (Zhai et al., 2018).

2.4. Japanese Language

Japanese is an agglutinative SOV language with topic-comment sentence structure. Phrases are exclusively head-final, and compound sentences are strictly left-branching. The most noticeable characteristics for language students are the missing articles, no distinction between singular and plural, no gender, no conjugation for person, a complex system of honorifics, and a high level of ambiguity, e.g. by omitting the subject or using zero anaphora. There exist many excellent reference grammars, e.g. Kamermans (2010); Kaiser et al. (2013), and a lot of research activity on Japanese linguistics (see Hasegawa, 2015, 2018).

One of the main obstacles for getting proficient in Japanese is the complex writing system (see Matsumoto, 2007; Mori, 2014; Paxton, 2019). It uses a combination of logographic *kanji* and two syllabaries *hiragana* and *katakana*. Kanji are adopted Chinese characters, since 2010 Japanese students are required to learn 2,136 so-called *jōyō kanji* in primary and secondary school. Most kanji have more than one reading depending on the context.

The most important lexical resource for Japanese is the *Japanese Multilingual dictionary* (JMdict) (Breen, 2004), which can be searched online in combination with many other lexical resources via the *Online Japanese Dictionary Service* (WWWJDIC)[10].

Another very useful online service is *Honyaku Star*[11]. It references numerous dictionaries and corpora and shows translations in context. Honyaku Star includes currently over 2 million translations. Japanese is also part of the *Open Multilingual Wordnet* (OMW) (Bond and Paik, 2012)[12], which makes it possible to assign Japanese words to English synsets. OMW is easily accessible via the *NLTK* toolkit[13].

The most prolific linguistic tool for Japanese is certainly the *CaboCha* dependency parser (Kudo and Matsumoto, 2002), which includes the *MeCab* part-of-speech and morphological analyzer (Kudo et al., 2004). More recently, trained pipelines have been added to the popular natural language toolkit *SpaCy*[14], another similar solution is *UniDic2UD*[15].

[10] http://wwwjdic.se/
[11] http://honyakustar.com/
[12] https://omwn.org/
[13] https://www.nltk.org/
[14] https://spacy.io/models/ja
[15] https://github.com/KoichiYasuoka/UniDic2UD

3. User Interface

In this section, we introduce our meaning representation by providing examples of the visual rendering in the user interface. The technical details are addressed later in Sect. 4.

The choices leading to the current user interface design are mainly based on practical experience and user feedback from previous research on meaning representation (Wloka and Winiwarter, 2021a), kanji acquisition (Wloka and Winiwarter, 2021b), and multimodal analogies (Winiwarter and Wloka, 2023) as part of a Web-based Japanese language learning environment. In all three cited publications we exclusively used images from Wikimedia Commons which are embedded in Wikipedia pages. The decision to restrict ourselves to this image source was mainly motivated by licensing issues but also by the valuable contextual semantic information accessible through the links to the original Wikipedia page(s).

The idea of using emojis to represent roles originated from our previous research work on using kanji within educational strategic games to foster incidental learning (Winiwarter, 2017). In recent implementations we increasingly relied on emojis to communicate additional gameplay information. This way we successfully assisted the user in focusing on kanji by eliminating other textual elements from the display. For a recent survey of research on emojis we refer to Bai et al. (2019).

As running example text for showcasing our user interface design, we use "From the Earth to the Moon"[16], an 1865 novel by Jules Verne, together with its English[17] and Japanese[18] translations. Figure 1 shows the AMR of the first sentence of the English version displayed in the *AMR Editor*[19]. The corresponding VOLARE representation is shown in Fig. 2. Each concept is visualized by a representative image. Whenever the user hovers over a concept, a tooltip with the concept identifier is displayed. For the ease of the reader, we have added the tooltip texts to Fig. 2 and the following user interface figures. A click on a concept shows an enlarged version of the image as well as the concept identifier and gloss in the bottom right corner of the screen.

In addition to *PropBank rolesets*, color-coded in teal, we can also observe *Wikipedia page titles* as concept identifiers shown in magenta:

- club → `Club_(organization)`,

[16] https://fr.wikisource.org/wiki/De_la_Terre_à_la_Lune
[17] https://en.wikisource.org/wiki/From_the_Earth_to_the_Moon
[18] https://ja.wikisource.org/wiki/地球から月へ
[19] https://amr.isi.edu/editor.html

Sentence: During the War of the Rebellion, a new and influential club was established in the city of Baltimore in the State of Maryland.

```
(e / establish-01
    :ARG1 (c / club
        :ARG1-of (n / new-01)
        :ARG0-of (i2 / influence-01))
    :location (c2 / city :name (n2 / name :op1 "Baltimore")
        :location (s / state :name (n3 / name :op1 "Maryland")))
    :time (w / war :name (n4 / name :op1 "War" :op2 "of" :op3 "the" :op4 "Rebellion")))
```

Figure 1: Example AMR.

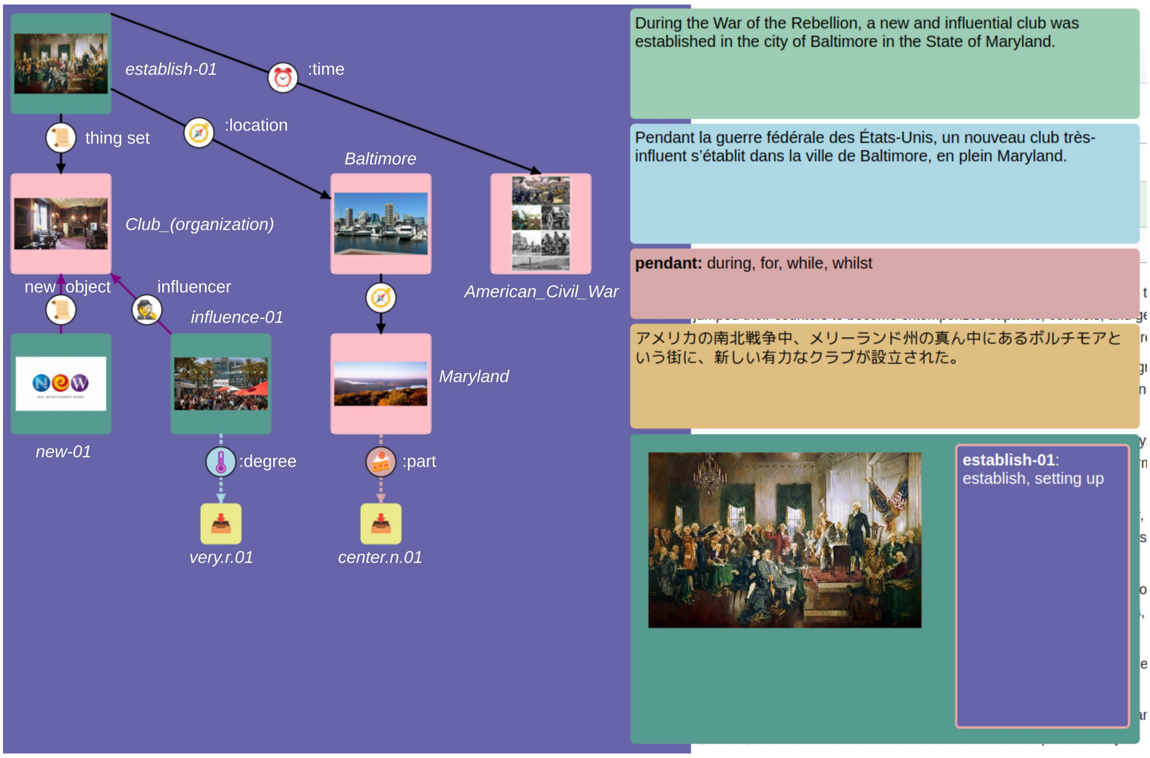

Figure 2: Example of user interface.

- Baltimore → `Baltimore`,
- Maryland → `Maryland`,
- War of the Rebellion → `American_Civil_War`.

As can be seen, this leads to a more concise and uncluttered representation of named entities.

AMR roles are visualized using *emojis*, the original roles are available as tooltips. For core arguments, we choose the emoji according to the role indicated in the PropBank frame file and show the gloss as tooltip. Inverse roles are indicated by inverted arrows and the line color violet. On the right side in Fig. 2 we display the sentence in English, French, and Japanese. By clicking on a word in either the French or Japanese version, lexical information with English glosses can be displayed, e.g. for the French word "pendant" in this case.

One important extension of AMR are the two *translation deviation annotations* (TDAs) in Fig. 2. The left one represents the word "très" in the French original, which is missing in the English translation. It is mapped to the *WordNet synset* `very.r.01` and therefore color-coded in yellow.

To keep the representation manageable, we only show an emoji, however, the detailed information can still be displayed in the bottom right corner. The relation for the role `:degree` (🌡️) is drawn as dashed line in the color cyan to indicate the language French.

In the same way, we map the two additional expressions "en plein" and "真ん中" to the synset `center.n.01` and link it to `Maryland` by the role `:part` (🍰). This line is drawn in purple to indicate

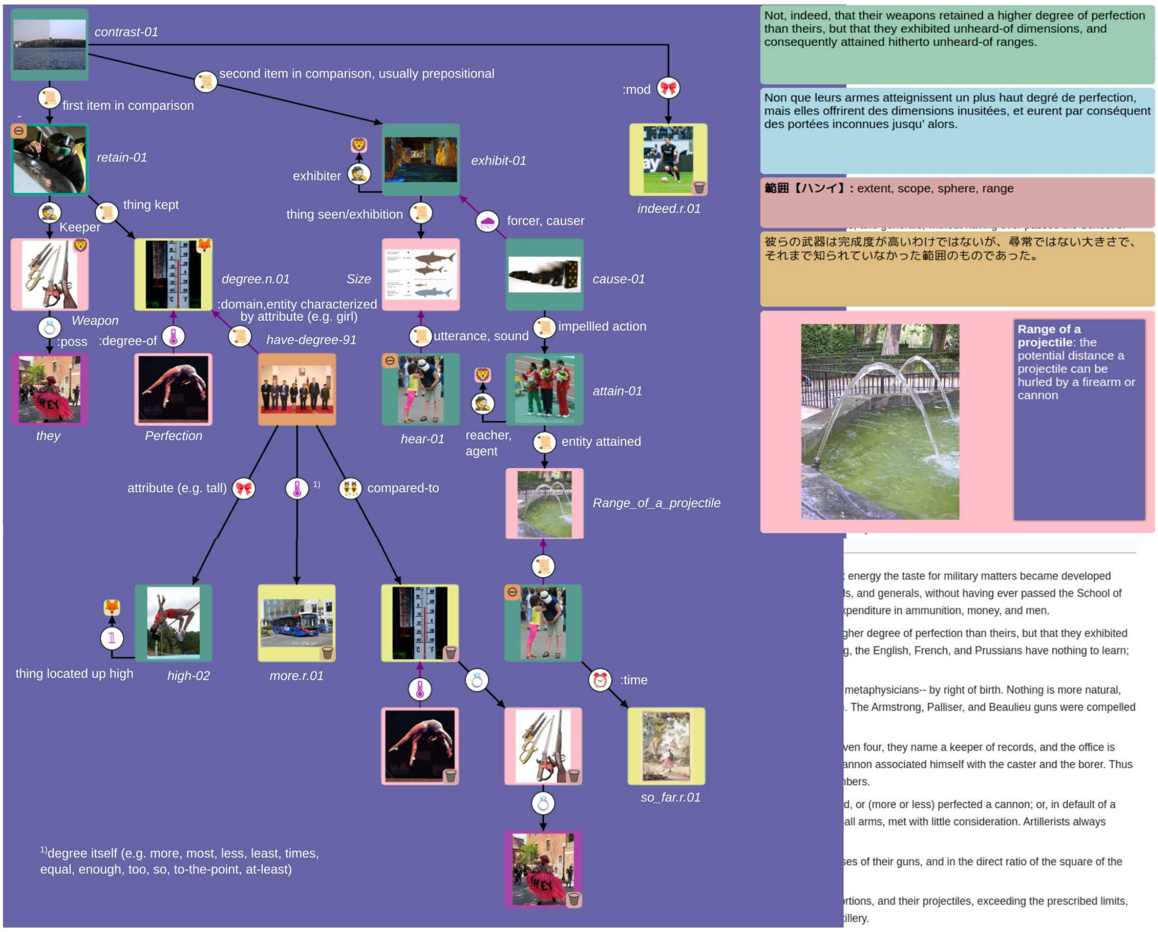

Figure 3: Example of user interface with co-reference.

Sentence: Not, indeed, that their weapons retained a higher degree of perfection than theirs, but that they exhibited unheard-of dimensions, and consequently attained hitherto unheard-of ranges.

```
(c / contrast-01
    :ARG1 (r / retain-01 :polarity -
        :ARG0 (w / weapon
            :poss (t / they))
        :ARG1 (d / degree
            :degree-of (p / perfection)
            :ARG1-of (h / have-degree-91
                :ARG2 (h2 / high-02
                    :ARG1 d)
                :ARG3 (m / more)
                :ARG4 (d2 / degree
                    :degree-of (p2 / perfection)
                    :poss (w2 / weapon
                        :poss (t2 / they))))))
    :ARG2 (e / exhibit-01
        :ARG0 w
        :ARG1 (d3 / dimension
            :ARG1-of (h3 / hear-01 :polarity -))
        :ARG0-of (c2 / cause-01
            :ARG1 (a / attain-01
                :ARG0 w
                :ARG1 (r2 / range
                    :ARG1-of (h4 / hear-01 :polarity -
                        :time (h5 / hitherto))))))
    :mod (i2 / indeed))
```

Figure 4: Example AMR with co-reference.

both French and Japanese (just Japanese would be orange).

In VOLARE, we only use variables for *co-reference*. Figure 3 and Fig. 4 give an example. We use animal emojis as variables (e.g. 🦁, 🦊). To avoid overloading the display, we draw the variable at the top right corner of the concept and duplicate it as target of co-referential links.

As can be seen, we visualize *negation* of concepts with a ⊖ symbol in the top left corner. Additional concept types used in this example are *Wikidata lexeme senses* (violet) and *special AMR frames* (orange).

Finally, the *wastebaskets* (🗑️) in the bottom right corners are translation deviation annotations that indicate concepts that are missing in the respective languages. The lexical information shown in Fig. 3 is for the Japanese word "範囲". It provides the pronunciation【ハンイ】in katakana, which corresponds to "han'i", and the English glosses "extent, scope, sphere, range".

One common case of translation deviation is the *substitution* of one or several concepts in the source language by alternative concepts in the target language. An example of a simple concept substitution is shown in Fig. 5. The concept displayed in the image for the English text is `Pocket_pistol`, which is an exact translation of the French original “pistolets de poche”. However, in Japanese it is translated as 拳銃, which can be mapped to the concept `Pistol`. This deviation is indicated by the symbol ⇄ in the bottom right corner. By clicking on the symbol, the user can inspect the detailed information about this concept.

Figure 5: Example of simple substitution.

A more complex situation is depicted in Fig. 6. The corresponding AMR snippet is:

```
(b / become-01
 :ARG1 (t / taste
   :topic (m / matter
     :topic (m2 / military)))
 ...
```

In this case, the original French expression for “the taste for military matters” is “ l’instinct militaire”. This is indicated by a substitute concept `Instinct` with the relation `:ARG1` (`entity changing`) from `become-01` and a relation `:mod` to `Military`. Since the Japanese text offers a precise translation of the French original, the deviation from the English version concerns both languages.

4. Implementation

In this section we will describe some details about the implementation of our Web-based annotation environment. We first provide a top-level overview of the system architecture, before we zoom in on the three subtasks `Preprocess`, `Annotate`, and `Customize` in separate subsections.

Figure 7 highlights the main components of our architecture. The users can access the server through a Web browser by using augmented browsing enabled through *Chrome extension APIs*[20], and the *jQuery*[21] and *jQuery UI*[22] libraries.

If a student loads a new Wikisource document in one of the three languages English, French, or Japanese, it is automatically analyzed and segmented into individual sentences. Each sentence is

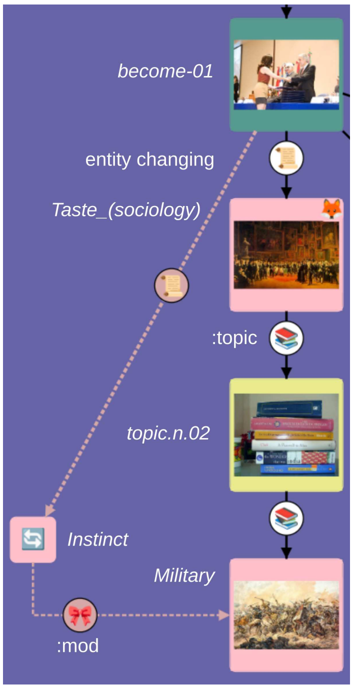

Figure 6: Example of complex substitution.

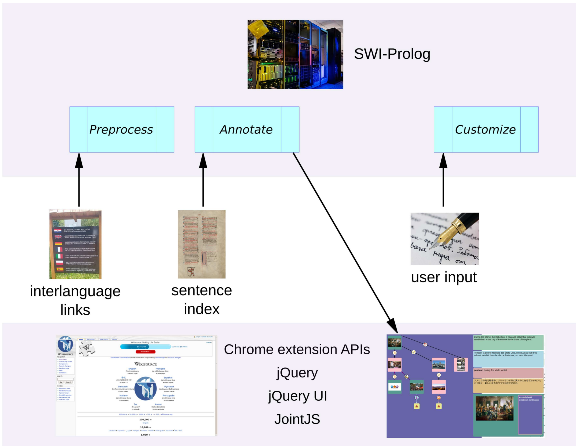

Figure 7: System architecture.

augmented with an event handler so that whenever a student then clicks on a sentence, it is transferred to the server.

If the Web document is a new text, we use the *interlanguage links* to retrieve the other two language versions and `Preprocess` the resulting document triplet. For existing texts, we use the *sentence index* to identify the selected sentence and `Annotate` it to produce the VOLARE annotation, which is sent back to the user. The rendering at the Web client is realized using the JavaScript interactive diagramming library *JointJS*[23] based on SVG. Among the many available diagramming solutions, we chose JointJS mainly because it is open

[20] https://developer.chrome.com/docs/extensions/reference/api
[21] https://jquery.com/
[22] https://jqueryui.com/
[23] https://www.jointjs.com/

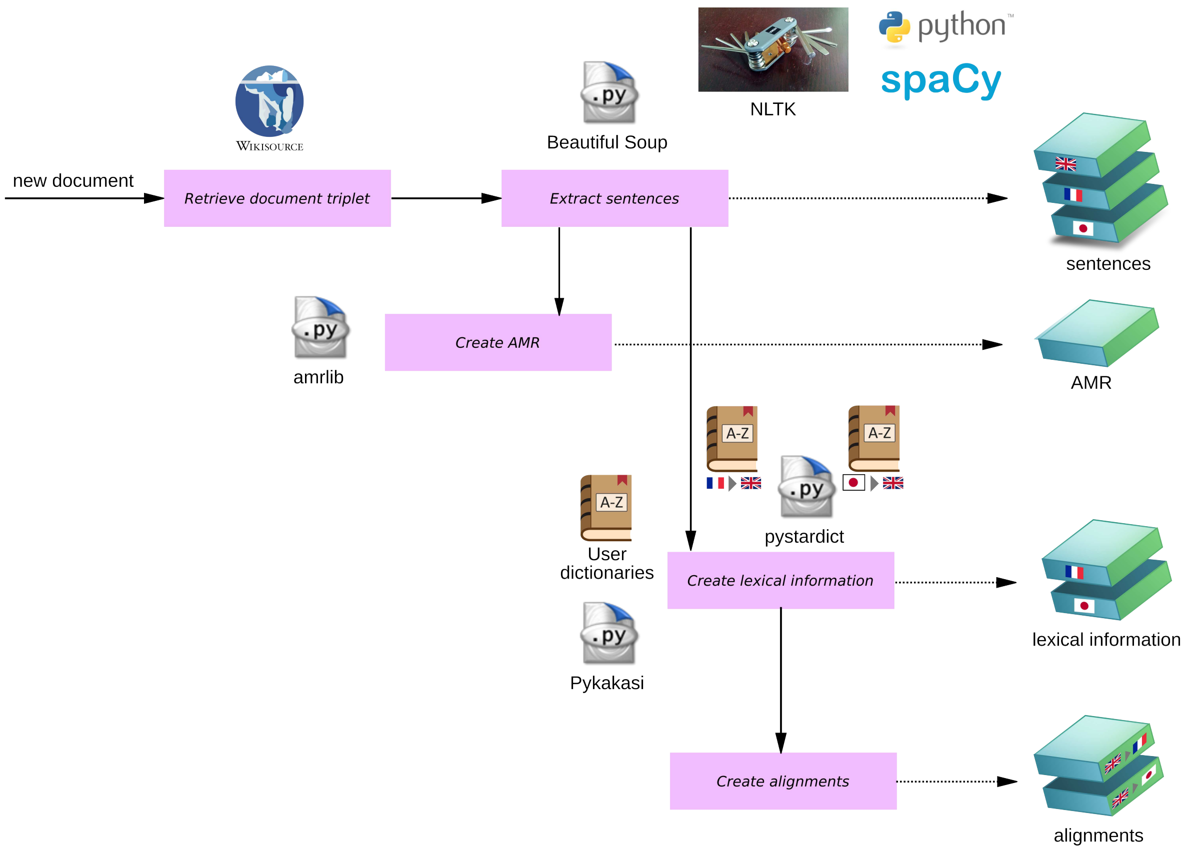

Figure 8: Subtask Preprocess.

source software[24], feature-rich, compatible with our augmented browsing scenario, and offers excellent documentation with many tutorials and demos.

The interactive diagramming library makes it possible to freely edit any aspect of the presented annotation. Each *user input* is sent to the server and leads to an update of the personal knowledge base used to `Customize` the annotation.

The annotation server is implemented in *SWI-Prolog*[25] (Wielemaker et al., 2012), which is not only an obvious choice for natural language processing tasks but also provides a scalable Web server solution (Wielemaker et al., 2008) and libraries for efficiently handling RDF and XML files.

4.1. Subtask Preprocess

Figure 8 gives an overview of the individual steps to preprocess a *new document*. In general, we have mainly written *Python* scripts for that purpose. In addition, we make use of the popular NLP toolkits *NLTK* and *SpaCy*. The latter provides trained models and pipelines for several languages including English, French, and Japanese.

The first step for a new English, French, or Japanese document is to `Retrieve the document triplet` by downloading the other two language versions from *Wikisource*. The three documents are then parsed by using the Python library *Beautiful Soup*[26] to `Extract` the individual `sentences`. Based on the resulting sentence collection, we can start to `Create` the `AMR` and the `lexical information`. These two steps can therefore be performed in parallel. We use the Python library *amrlib*[27] to produce the AMR, which is available as SpaCy extension[28].

To create the lexical information, in particular English glosses for French and Japanese words and expressions, we use the Python library *pystardict*[29] to access the French-English and the Japanese-English *StarDict dictionaries*[30]. The latter is based on the popular JMdict dictionary. In addition, we look up the personal *user dictionaries* (see Sect. 4.3) to include customized entries.

[24] https://sourceforge.net/projects/jointjs/
[25] https://www.swi-prolog.org
[26] https://www.crummy.com/software/BeautifulSoup/
[27] https://github.com/bjascob/amrlib
[28] https://spacy.io/universe/project/amrlib
[29] https://github.com/lig/pystardict
[30] https://stardict-4.sourceforge.net/

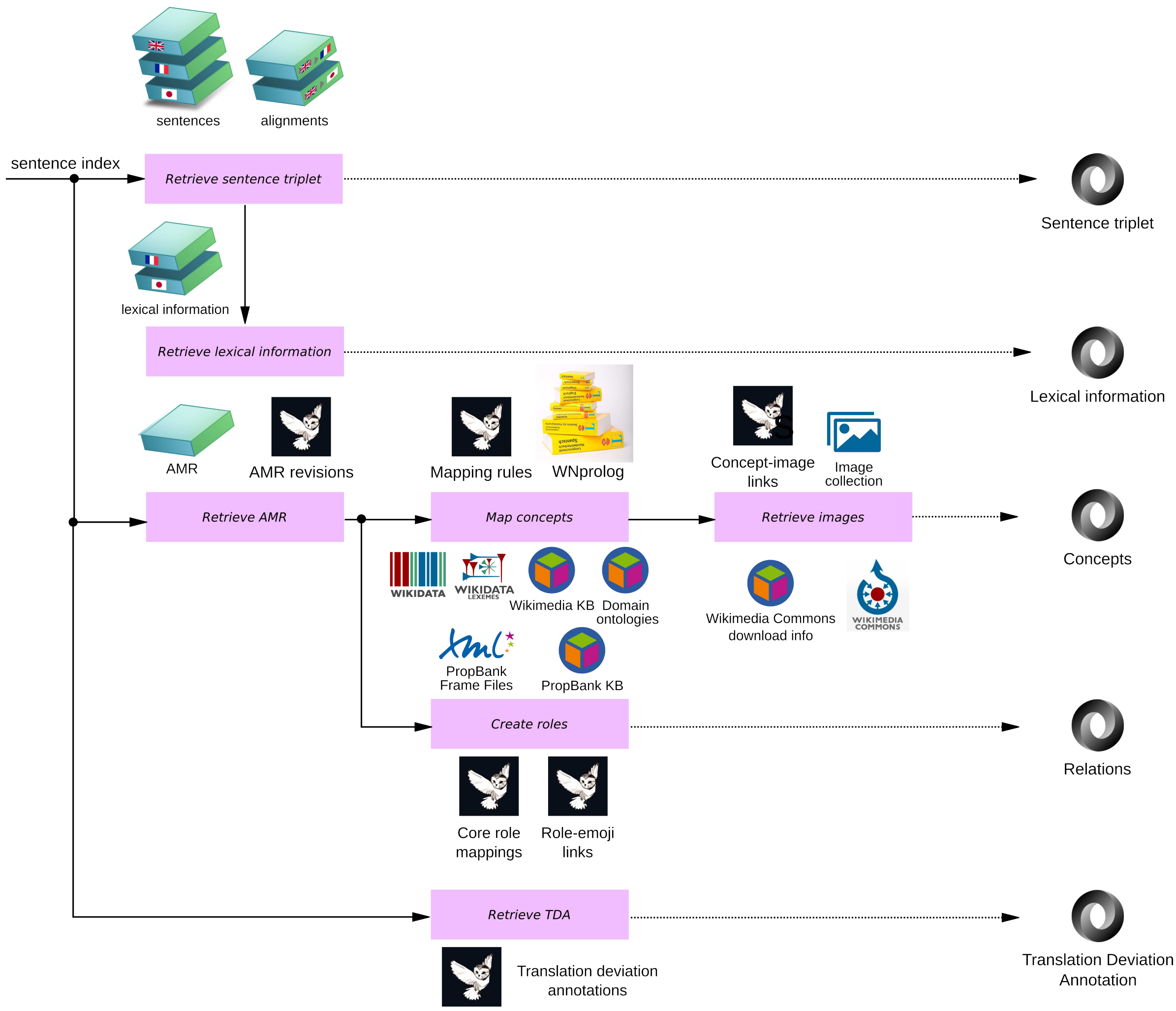

Figure 9: Subtask Annotate.

Finally, we use the Python library *Pykakasi*[31] based on the transliteration tool *kakasi*[32] to add the correct pronunciation to Japanese words with kanji characters. The last preprocessing step is to `Create` the bilingual `alignments` between the language pairs English-French and English-Japanese. Based on the lexical information we calculate a similarity value to align sentences depending on the comparison with threshold values. Although $1:1$ and $1:2$ are the most common cases, we are able to handle all theoretically possible $0 \ldots n : 0 \ldots m$ patterns.

For performance reasons, we first execute the necessary preprossing steps for the sentence that the user wants to inspect so that the annotation subtask can start without any significant delay for the user. This means that the remainder of the document is analyzed as background process while the user can already interact with the annotation at the client. We also keep all the required SpaCy models and pipelines preloaded to save valuable initialization time.

[31] https://pypi.org/project/pykakasi/

[32] http://kakasi.namazu.org/

4.2. Subtask Annotate

The details of this subtask are depicted in Fig. 9. Based on the resources created in the previous subsection, we generate *JSON* objects for the individual components of the annotation. We apply various types of customization data provided by the user (see Sect. 4.3) in several processing steps.

To produce the list of concepts for rendering in VOLARE, we first `Retrieve` the `AMR` and correct it according to the *AMR revisions*. The next step is to `Map` the `concepts` in the AMR to uniquely identifiable sense definitions. For *PropBank rolesets*, we only have to add glosses from our *PropBank KB*, which we extracted from the *PropBank Frame Files*, which are in *XML* format. All other concepts, we first try to match with *Wikipedia pages* and the corresponding *Wikidata items*. The necessary information is stored in our *Wikimedia KB*, which we generated with the use of *DBpedia* (Lehmann et al.,

2015) datasets. Next, we attempt to find compatible *WordNet* (Princeton University, 2012) synsets via *WNprolog*. Since existing word sense disambiguation solutions such as *pywsd*[33] produced unsatisfactory results, we have developed our own model, which we continuously improve based on user input stored as *mapping rules*. Any remaining concepts are mapped to *Wikidata Lexemes*. Finally, the user can add *domain ontologies* to this collection of ontological resources.

The last missing task for creating the concept description is to `Retrieve` the representative `images`. We store for all Wikipedia pages in our Wikimedia KB the *Wikimedia Commons download info* so that we can download any images for concepts that are accessed for the first time. WordNet synsets are mapped to images in our *image collection* according to the available synset associations. The users can add their own *concept-image links* to personalize their visualization.

For the relation description, one important task is to `Create` the `role` for display by mapping core roles to thematic roles according to the defined *core role mappings*, and to translate the role names to emojis by following the *role-emoji links*. Both can be freely adjusted to suit the preferences of the user. The role glosses for the tooltips are retrieved from the PropBank KB. Finally, if there exists any *translation deviation annotation* for this sentence, we `Retrieve` the `TDA` and add it to the JSON array to complete the VOLARE annotation.

4.3. Subtask Customize

All the processing steps in Fig. 8 and Fig. 9 are carried out fully automatically. However, in many cases there is still room for improvement or the desire for adjustments to better adapt the annotation results to individual needs and preferences.

Since we have already covered many aspects of customization in the previous subsections, we can keep the presentation here brief and just summarize the numerous possibilities offered to the user. Enabled by the interactive diagramming functionality of JointJS, the users can freely edit any VOLARE annotation to fine-tune it to better suit their personal preferences. Any user input is sent to the server where it is processed and leads to an update of the affected resources.

Any changes to the original AMR are stored in the *AMR revisions* and used to correct the AMR if the sentence annotation is displayed again in future. In the same way, any *translation deviation annotation* is saved and can be reviewed at a later time. The users can also change concepts in the annotation leading to an update of the *mapping rules*; they can add *domain ontologies*; and add new images to the *image collection* or use existing images to represent a concept, which changes the *concept-image link*. Similarly, they can choose different roles for core arguments, which updates the *core role mappings*, as well as other emojis, which results in an actualized *role-emoji link*. Finally, the users can also improve the display of the lexical information by adding new entries to their *user dictionaries*.

An important aspect is that in classroom or company environments, the individual customization data can be collected, analyzed, and consolidated to create integrated resources at the organizational unit level.

5. Conclusion

In this paper, we have presented a Web-based annotation environment, which extends AMR with visual and ontological elements. The addition of TDAs enables the comparative analysis of the different language versions of a document.

We will evaluate our use case implementation with several volunteer professional translators and interpreters. Based on the achieved results including user feedback concerning functionality and usability aspects, we will further improve our system and draw up annotation guidelines for the users.

Interesting aspects for the evaluation will be the impact of adding representative images on the quality of the translation process, a detailed analysis of the customizations performed by the users, and experiments from a cognitive linguistic perspective, e.g. regarding the question whether a multilingual speaker has a single cross-lingual visual representation of a concept or different visual representations depending on the language currently used.

We have also already planned classroom scenarios in university courses to investigate the educational benefits of our environment and the challenges/opportunities that arise from aggregating and harmonizing the individual customization data.

The extension to other languages is straightforward as long as StarDict dictionaries and SpaCy models and pipelines exist. We can also easily accommodate more than three languages, however, this will require some filtering regarding the TDAs, otherwise the display will become too cluttered.

The main challenge for future work is to switch the foundation of VOLARE to UMR, which is also much better suited for multilingual annotations. The incorporation of the UMR document-level representation will make it possible to model intersentential dependencies, which will at the same time lead to new requirements for our environment and TDAs. We have already started with some first considerations and preparatory work. Thus, we eagerly await the availability of UMR parsers to begin with real experiments and implementation work.

[33] https://github.com/alvations/pywsd

6. Bibliographical References

Omri Abend and Ari Rappoport. 2017. The state of the art in semantic representation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 77–89, Vancouver, Canada. Association for Computational Linguistics.

Qiyu Bai et al. 2019. A systematic review of emoji: Current research and future perspectives. *Frontiers in Psychology*, 10.

Laura Banarescu et al. 2013. Abstract Meaning Representation for sembanking. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse*, pages 178–186, Sofia, Bulgaria. Association for Computational Linguistics.

Michele Bevilacqua, Rexhina Blloshmi, and Roberto Navigli. 2021. One SPRING to rule them both: Symmetric AMR semantic parsing and generation without a complex pipeline. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2021)*, pages 12564–12573.

Arne Bewersdorff et al. 2024. Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. arXiv:2401.00832 [cs.AI].

Francis Bond and Kyonghee Paik. 2012. A survey of wordnets and their licenses. In *Proceedings of the 6th Global WordNet Conference (GWC 2012)*, pages 64–71.

Francis Bond et al. 2009. Enhancing the Japanese WordNet. In *Proceedings of the 7th Workshop on Asian Language Resources (ALR7)*, pages 1–8, Suntec, Singapore. Association for Computational Linguistics.

James Breen. 2004. JMdict: A Japanese-Multilingual dictionary. In *Proceedings of the Workshop on Multilingual Linguistic Resources*, pages 71–79. Association for Computational Linguistics.

Jon Z. Cai et al. 2023. CAMRA: Copilot for AMR annotation. arXiv:2311.10928 [cs.CL].

Agostina Calabrese, Michele Bevilacqua, and Roberto Navigli. 2020. Fatality killed the cat or: BabelPic, a multimodal dataset for non-concrete concepts. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4680–4686, Online. Association for Computational Linguistics.

Khyathi Raghavi Chandu, Yonatan Bisk, and Alan W Black. 2021. Grounding 'grounding' in NLP. arXiv 2106.02192 [cs.CL].

Pin-Er Chen et al. 2023. Exploring affordance and situated meaning in image captions: A multimodal analysis. In *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation*, pages 293–302, Hong Kong, China. Association for Computational Linguistics.

William Croft. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press.

William Croft. 2002. *Typology and Universals*, 2nd edition. Cambridge Textbooks in Linguistics. Cambridge University Press.

William Croft. 2022. *Morphosyntax: Constructions of the World's Languages*. Cambridge Textbooks in Linguistics. Cambridge University Press.

William Croft. 2023. Word classes in Radical Construction Grammar. In *The Oxford Handbook of Word Classes*. Oxford University Press.

Dun Deng and Nianwen Xue. 2017. Translation divergences in Chinese-English machine translation: An empirical investigation. *Computational Linguistics*, 43(3):521–565.

Jia Deng et al. 2009. ImageNet: A Large-Scale Hierarchical Image Database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Miami, FL.

Dagmar Divjak, Petar Milin, and Srdan Medimorec. 2020. Construal in language: A visual-world approach to the effects of linguistic alternations on event perception and conception. *Cognitive Linguistics*, 31(1):37–72.

Lydia Feng et al. 2023. Widely interpretable semantic representation: Frameless meaning representation for broader applicability. arXiv:2309.06460 [cs.CL].

Mark Fishel, Ondřej Bojar, and Maja Popović. 2012. Terra: a collection of translation error-annotated corpora. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 7–14, Istanbul, Turkey. European Language Resources Association (ELRA).

Jens E. L. Van Gysel et al. 2021. Designing a Uniform Meaning Representation for natural language processing. *KI – Künstliche Intelligenz*, 35:343–360.

Christopher Hart and Javier Marmol Queralto. 2021. What can cognitive linguistics tell us about language-image relations? A multidimensional approach to intersemiotic convergence in multimodal texts. *Cognitive Linguistics*, 32(4):529–562.

Yoko Hasegawa. 2015. *Japanese: A Linguistic Introduction*. Cambridge University Press, Cambridge, UK.

Yoko Hasegawa, editor. 2018. *The Cambridge Handbook of Japanese Linguistics*. Cambridge Handbooks in Language and Linguistics. Cambridge University Press.

Stefan Kaiser et al. 2013. *Japanese: A Comprehensive Grammar*, 2nd edition. Routledge, London and New York.

Michiel Kamermans. 2010. *An Introduction to Japanese – Syntax, Grammar & Language*. SJGR Publishing, Rotterdam, The Netherlands.

Taku Kudo and Yuji Matsumoto. 2002. Japanese dependency analysis using cascaded chunking. In *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002 (COLING 2002 Post-Conference Workshops)*, pages 63–69.

Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. 2004. Applying conditional random fields to Japanese morphological analysis. In *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, pages 230–237, Barcelona, Spain. Association for Computational Linguistics.

Ronald W. Langacker. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford University Press.

Young-Suk Lee et al. 2022. Maximum Bayes Smatch ensemble distillation for AMR parsing. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5379–5392, Seattle, United States. Association for Computational Linguistics.

Jens Lehmann et al. 2015. DBpedia – a large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web*, 6(2):167–195.

Paul Lerner, Olivier Ferret, and Camille Guinaudeau. 2024. Cross-modal retrieval for knowledge-based visual question answering. arXiv:2401.05736 [cs.CL].

Hiroshi Matsumoto. 2007. Peak learning experiences and language learning: A study of American learners of Japanese. *Language Culture and Curriculum - LANG CULT CURRIC*, 20:195–208.

Yoshiko Mori. 2014. Review of recent research on kanji processing, learning, and instruction. *Japanese Language and Literature*, 48(2):403–439.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TACL)*, 2:231–244.

Roberto Navigli et al. 2021. Ten years of BabelNet: A survey. In *Proceedings of IJCAI 2021*, pages 4559–4567.

Simon Paxton. 2019. Kanji matters in a multilingual Japan. *The Journal of Rikkyo University Language Center*, 42:29–41.

James Pustejovsky et al. 2016. The development of multimodal lexical resources. In *Proceedings of the Workshop on Grammar and Lexicon: interactions and interfaces (GramLex)*, pages 41–47, Osaka, Japan. The COLING 2016 Organizing Committee.

Andrew Rothwell et al. 2023. *Translation Tools and Technologies*, 1st edition. Routledge, London and New York.

Shejla Tahiri. 2020. The impact of pictures on second language acquisition. *SEEU Review*, 15(2):126–135.

Jan Wielemaker, Zhisheng Huang, and Lourens Van Der Meij. 2008. SWI-Prolog and the Web. *Theory and Practice of Logic Programming*, 8(3):363–392.

Jan Wielemaker et al. 2012. SWI-Prolog. *Theory and Practice of Logic Programming*, 12(1-2):67–96.

Werner Winiwarter. 2017. KANGAROO – the kanji game room. In *Proceedings of the 19th International Conference on Information Integration and Web-based Applications & Services*, pages 535–542, New York. ACM.

Werner Winiwarter and Bartholomäus Wloka. 2022. VISCOSE – a kanji dictionary enriched with VISual, COmpositional, and SEmantic information. In *7th Workshop on Cognitive Aspects of the Lexicon (CogALex-VII)*, pages 68–77.

Werner Winiwarter and Bartholomäus Wloka. 2023. CLE-UMA – a creative learning environment using multimodal analogies. In *The 17th International Conference on Knowledge, Information and Creativity Support Systems*, IIAI Letters on Informatics and Interdisciplinary Research.

Bartholomäus Wloka, Yves Lepage, and Werner Winiwarter. 2022. WAPITI – web-based assignment preparation and instruction tool for interpreters. In *24th International Conference on Information Integration and Web Intelligence (iiWAS2022)*, pages 295–306.

Bartholomäus Wloka and Werner Winiwarter. 2021a. AAA4LLL – Acquisition, Annotation, Augmentation for Lively Language Learning. In *3rd Conference on Language, Data and Knowledge (LDK2021)*.

Bartholomäus Wloka and Werner Winiwarter. 2021b. DARE – a comprehensive methodology for mastering kanji. In *23rd International Conference on Information Integration and Web Intelligence (iiWAS2021)*, pages 427–435.

Yuming Zhai, Aurélien Max, and Anne Vilnat. 2018. Construction of a multilingual corpus annotated with translation relations. In *Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing*, pages 102–111, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

7. Language Resource References

Princeton University. 2012. *WordNet 3.1*. Princeton University, ISLRN 379-473-059-273-1.