Manzanilla: An Image Annotation Tool for TKB Building

Arianne Reimerink, Pilar León-Araúz
University of Granada
Department of Translation and Interpreting, C/ Buensuceso, 11, 18008 Granada (Spain)
arianne@ugr.es, pleon@ugr.es

Abstract

Much has been written regarding the importance of combining visual and textual information to enhance knowledge acquisition (Paivio, 1971, 1986; Mayer & Anderson, 1992). However, the combination of images and text still needs further analysis (Faber, 2012; Prieto, 2008; Prieto & Faber, 2012). An in-depth analysis of the features of images provides the means to develop selection criteria for specific representation purposes. The combination of conceptual content, image type based on morphological characteristics, and functional criteria can be used to enhance the selection and annotation of images that explicitly focus on the conceptual propositions that best define concepts in a knowledge base. Manzanilla is an image annotation tool specifically created for EcoLexicon, a multilingual and multimodal terminological knowledge base (TKB) on the environment. It is powered by Camomile (Poignant et al., 2016) according to the selection and annotation criteria resulting from ten years of research on multimodality within the framework of Frame-Based Terminology (FBT; Faber, León-Araúz & Reimerink, 2014). The tool was created to enhance the consistency of knowledge representation through images with the conceptual knowledge in EcoLexicon and to improve image reusability.

Keywords: image annotation, multimodal knowledge representation, EcoLexicon, Camomile

1. Introduction

Manzanilla is an image annotation tool specifically created for EcoLexicon, a multilingual and multimodal terminological knowledge base (TKB) on the environment[1]. It was developed with Camomile (Collaborative Annotation of multi-MOdal, multI-Lingual and multi-mEdia documents; Poignant et al., 2016)[2] according to the selection and annotation criteria resulting from ten years of research on multimodality within the framework of Frame-Based Terminology (FBT; Faber, León-Araúz & Reimerink, 2014). The tool was created to enhance the consistency of knowledge representation through images with the conceptual knowledge in EcoLexicon and to improve image reusability.

Currently, images are stored in the TKB in association with concept entries according to the semantic content described in their definition, and are thus regarded as a whole and only linked to the concept itself. Other knowledge bases, such as BabelNet, the automatically constructed multilingual encyclopedic dictionary and semantic network (Navigli & Ponzetto, 2012), also uses this approach. However, regarding images as a whole does not allow for a more fine-grained annotation where the semantic relations between different concepts represented in an image are made explicit. Our new approach is that images should not be stored in the TKB as the representation of a concept, but as the representation of a set of conceptual propositions (concept-relation-concept triples) more in line with the conceptual structure of EcoLexicon. Therefore, images must be annotated according to semantic and morphological information and stored in a separate repository. Since each image activates several propositions and each proposition can be activated by different concepts, one image can then be linked to several concept entries. This would enhance the reusability of images, improve the consistency of the TKB and avoid duplicating workload (Reimerink, León-Araúz & Faber, 2016; León-Araúz & Reimerink, 2016).

In Section 2, we explain how images have been selected and included in EcoLexicon up to now. In Section 3, a summary of our research into image selection and annotation criteria is given. Then, in Section 4, the tool is explained in detail. Finally, in Section 5 some conclusions are drawn and future work is addressed.

2. Images in EcoLexicon

The knowledge contained in EcoLexicon is largely based on information extracted from a specialized domain corpus that was compiled for this specific purpose (Faber, León-Araúz & Reimerink, 2014; Faber, León-Araúz & Prieto, 2009). This conceptual knowledge is represented through semantic networks based on conceptual propositions and definitions based on these networks. To further enrich conceptual description, a visual corpus was also compiled. As stated above, each concept entry has several images selected according to the semantic content of a concept's definition.

The definitions in EcoLexicon are based on templates that define category membership and describe the basic conceptual propositions in which the concept participates. In this way, definitions have a uniform structure that directly refers to and evokes the underlying conceptual structure of the domain, represented in the semantic networks.

For example, for the definition of WATER EROSION, the template includes the four basic relations of all natural processes: *is_a*, *has_agent*, *affects* and *has_result*. For the selection of images, the basic conceptual propositions in the definitional template are used to select images which contain the same information to reinforce knowledge acquisition (Faber et al., 2007). Figure 1 shows one of the images included in the conceptual entry of WATER EROSION to represent the relation *has_result*. The template also has an additional relation because it is a complex procedural concept, which can be divided into a sequence of steps: *has_phase*. Figure 2 was included in the entry to represent the conceptual proposition WATER EROSION *has_phase* WEATHERING (León-Araúz, Reimerink &

[1] EcoLexicon is freely accessible at ecolexicon.ugr.es.
[2] https://camomile.limsi.fr

Faber, 2013). Images are thus regarded as a whole and are only linked to the concept itself.

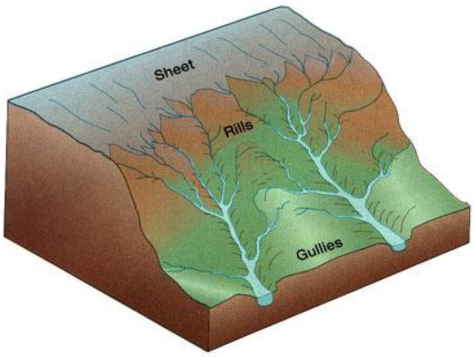

Figure 1: Image for SHEET/RILL/GULLY *result_of* WATER EROSION in EcoLexicon.

Figure 2: Image for WATER EROSION *has_phase* WEATHERING in EcoLexicon.

So far, we have shown how the same concept can be represented through different images, depending on perspective, or the semantic content highlighted (Faber et al., 2007; Reimerink, García de Quesada & Montero-Martínez, 2010). However, one and the same image may also work for the representation of other related concept entries (e.g. an entity and the process through which it was formed, a concept and its parts, etc.). Images should thus be further dissected according to the features they possess (i.e. image type and other morphological characteristics) and the knowledge they convey (i.e. semantic content). For example, many images show several concepts in a specific background where they establish different relations that can be explicitly labelled or inferred from previous knowledge. In this sense, we propose a different approach where images are stored in the TKB not as the representation of a concept, but as the representation of a set of conceptual propositions. Thus, images must be annotated according to semantic and morphological information and stored in a separate repository.

3. Image Selection Criteria

In previous research (Reimerink, Léon-Araúz & Faber 2016, León-Araúz & Reimerink, 2016), we have explained in detail how images convey conceptual knowledge through their morphological features, such as the use of colours, arrows, labels, etc., that we have called visual knowledge patterns (VKPs). In this section, we will summarize our findings related to the interaction between concept type, image type, and VKPs with a few examples. We use two functional criteria, referential similarity and dynamism, to analyse VKPs in images. Referential similarity refers to the degree to which an image resembles its referent in the real world. This similarity is measured on a continuum ranging from non-similar to totally identical. It goes without saying that a two-dimensional image can never be totally identical to its referent, but a colour photograph would have a high degree of referential similarity. Dynamism can also be measured on a continuum ranging from totally static to very dynamic. The results showed which VKPs and which degrees of referential similarity and dynamism are most characteristic of different types of images and how they are related to the conceptual propositions represented in each type.

It has also become clear that VKPs, such as arrows, labels, and colour-coding, are polysemic since the same pattern can be used for different purposes in the same way that textual knowledge patterns can also convey different conceptual relations (León-Araúz, Reimerink & Faber, 2009). Accordingly, the conceptual knowledge underlying VKPs can only be interpreted in the context of each image. Nevertheless, a certain combination of patterns, constrained by image and concept type, makes images more or less suitable for the representation of certain types of conceptual knowledge. An arrow, for example, can be used to connect a term to its representation in the image, thus this VKP does not necessarily transmit dynamism. However, when arrows appear in an image representing a process, they generally convey dynamism and go in the direction of the different phases of the process. The same is true for colours. In images with a high level of referential similarity, the colours in the image are the same or similar to those of the real world entity. In many cases, however, the function of the colours is not to realistically represent the concepts or its natural surroundings, but rather to differentiate closely related concepts in time or space.

For example, a GROYNE is a defence structure perpendicular to the coastline, which retards littoral drift and erosion. It can be made of stone, concrete or wood. The concept GROYNE is an entity and as it can be made of several materials, the proposition GROYNE *made_of* STONE/CONCRETE/WOOD will require more than one image.

Figure 3: Static image for GROYNE *made_of* WOOD[3].

[3] Source : http://blog.seamaidengemsjewellery.co.uk/

Figure 3 is a good example of an adequate image for the proposition GROYNE *made_of* WOOD. It is a static image and a colour photograph which provides a very high level of referential similarity. The same image can also be used for the proposition GROYNE *has_location* COAST. This relation is relevant when the location of a physical object is essential for its description. For instance, a groyne is not a groyne if it is not located on the coast.

Processes are generally described by the meronymic relations *phase_of* and *takes_place_in* because processes are composed of different stages and occur within a certain context. This is in direct contrast to physical objects (such as GROYNE), whose description is dominated by the relations *has_location* and *part_of*. Not surprisingly, processes are generally portrayed by flow charts that represent more than one relation. For example, Figure 4 is an image of the geological cycle, an extremely complex process, which shows both the *take_place_in* and *phase_of* relations. The concepts HARDENING, METAMORPHISM, MELTING, CRYSTALLIZATION, and INTRUSION *take_place_in* under the Earth's surface. At the same time, they are also *phases_of* GEOLOGICAL CYCLE. Figure 4 also conveys the *result_of* relation. This relation is relevant to either events or entities that are derived from other events. In this case, it shows SEDIMENTARY ROCK *result_of* HARDENING, METAMORPHIC ROCK *result_of* METAMORPHISM, etc. The representation of certain natural objects and events (e.g. sun, rain, clouds, magma, volcanic eruption, etc.) has a high degree of similarity to their referents in the real world.

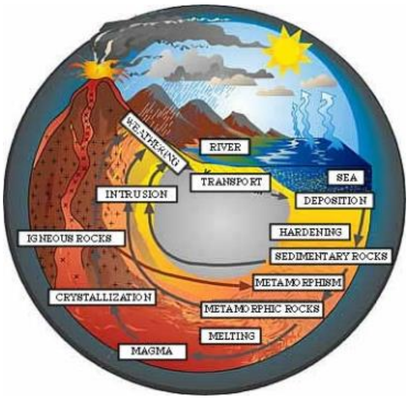

Figure 4: Dynamic image for GEOLOGICAL CYCLE[4].

However, other less well-known objects are labelled to explain where one type of geological formation ends and the other begins. This is the case of SEDIMENTARY ROCKS, METAMORPHIC ROCKS, MAGMA, and IGNEOUS ROCKS. Furthermore, the use of similar yet different colours heightens resemblance and, at the same time, delimits similar concepts or those that occur in connected locations. More specifically, WATER and SKY are different shades of blue, and there is a gradual colour change from yellow to red and dark brown to show how SEDIMENT becomes ROCK and then MAGMA.

Arrows add dynamism to images that portray how certain processes stem from others and how they affect one another. Thus, arrows as visual knowledge patterns (VKPs) most often convey meronymy in the case of entities (*part_of*) and (*phase_of*), and the *result_of* relation in the case of processes.

The findings of our research on the interrelations between concept type, image type, and VKPs have resulted in the following selection guidelines for researchers that work on images for EcoLexicon:

1. Use photographs for the *type_of*, *made_of*, and *has_location* relations of physical entities shown in their real-world environment.
2. Use drawings with labels and arrows for representing complex meronymic relations (*part_of*, *delimited_by*) or to differentiate between closely related concepts that are otherwise hard to differentiate without making reference to one another. Drawings are mostly fit to represent entities, but combinations of several drawings can be used to describe processes and their phases, especially if no flow chart is available.
3. Use flow charts for complex processes and non-hierarchical relations such as *causes* and *result_of*. The flow chart must show a high level of referential similarity for the background. It must use colour-contrast to differentiate between closely related concepts. It must also contain arrows to add dynamism and show the direction of the movement or even add textual explanations.

4. Manzanilla

Image annotation is often defined as the labelling of the semantic content of images with a set of keywords (Wenyin et al., 2001). However, "even though an image is worth a thousand of words, humans still possess the ability to summarize an image's contents using only one or two sentences. Similarly, humans may deem two images as semantically similar, even though the arrangement or even the presence of objects may vary dramatically" (Zitnick and Parikh, 2013).

One of the most famous annotated image sets that are available at present is ImageNet (Deng et al., 2009, Russakovsky et al., 2015), which is the largest annotated image set available and mostly consists of photographs that are annotated according to the hierarchical structure of WordNet synsets. This is done by automatically retrieving images from the internet through searches with WordNet synonyms, which are then verified for accuracy by humans through Amazon Mechanical Turk. BabelNet, the automatically constructed multilingual encyclopedic dictionary and semantic network (Navigli & Ponzetto, 2012) includes around 11 million images that are automatically retrieved from ImageNet and Wikipedia. The new EcoLexicon image repository we envision is different from these examples as ImageNet mostly includes photographs, whereas our repository considers other image types, such as drawings and flow charts which can represent more complicated specialized knowledge. Furthermore, the fine-grained annotation based on all the conceptual propositions contained in an image goes far beyond the synset annotation of ImageNet. In BabelNet, images are included in entries as a whole without further specifying the conceptual relationships contained in the images. This is also the case for

[4] Source: http://finstone.fi/engl/geology/

EcoLexicon at present, but that is exactly what we want to change with the annotation proposal in this paper. Apart from the above, ImageNet nor BabelNet provide much domain specific knowledge.

Manual image annotation, apart from inconsistent, can be very time-consuming. For this reason, in computer science automatic image annotation has been studied for some time now (Jeon, Lavrenko and Manmatha, 2003; Li and Wang, 2008). However, these studies mostly focus on photographs (Zitnick and Parikh, 2013 being one of the exceptions), objects and rather general concepts. Furthermore, they do not take into account the interaction of the semantic elements. In this sense, Mei et al. (2008) acknowledge that approaches to automatic image annotation do not usually guarantee good semantic coherence of the annotated words for each image, because they treat each word independently without considering the inherent semantic coherence among the words.

The neural network community has also addressed the problem of image classification focussing on automatic object recognition (Russakovsky et al. 2015) and even addressing fine-grained classification issues, such as recognizing subordinate-level categories (Xiao et al., 2015). The ImageNet Large Scale Visual Recognition Challenge (Russakovsky et al. 2015) has been run annually since 2010 and has achieved ground-breaking results in the area. Nevertheless, the neural network community mostly concentrates on object recognition and maybe the categorization of subtypes. To our best knowledge, the existing relation between different entities or processes in one and the same image are not taken into account in this field of research. Much of the effort goes into identifying the primary object in the foreground, while discarding the information contained in the background. For our purposes, the relation between entities and/or processes and the context or background (in which they are located or take place) are essential for representing the complex multidimensional knowledge of the environmental domain.

Unfortunately, given the specificity of the graphical information that an environmental TKB requires, where we intend to annotate the semantic relations between all concepts represented in each image, and the specialization of the field, at this point in time such procedures cannot be applied in our case.

Taking into account previous research results and the image selection guidelines, we developed Manzanilla. The tool was developed within the framework of Camomile with a step by step interface to facilitate annotation and ensure consistency.

4.1 Camomile

Image annotation forces us to clearly think about naming and categorization issues (Barriuso and Torralba, 2012), which are tasks that are not as straight-forward as they may seem. To facilitate the annotation process and enhance inter-annotator consistency, we needed a system with a flexible but strictly organized interface to label the different types of information related to each image in EcoLexicon. We opted for Camomile for its design because it is open source and flexible, as the user interfaces are specifically created for each use case (Poignant et al., 2016). Its collaborative annotation framework follows a client/server architecture, which facilitates the work of multiple users on consistent data sources. In the Camomile framework, resources are annotations, which are represented in JSON formats, stored in a MongoDB database. Based on the use case, four types of collections are developed: corpus, media, layers, and annotations. The corpus collection describes all available corpora. Each corpus contains a set of media and a set of layers. A medium corresponds to a multimedia resource (e.g., a video or audio file). A layer is composed of multiple annotations with the same type (e.g. one layer for manual annotations of speech turns or one layer for annotations of face tracks). An annotation is uniquely defined by a media fragment (e.g., a temporal segment) and attached data (e.g. the name of the current speaker) (idem: 1422).

For the design of the tool, we expressed our needs in the following guidelines for annotation of EcoLexicon images:
1. Annotate image type: photograph, drawing (including maps or diagrams), or flow chart.
2. Annotate all the concepts in EcoLexicon which are present in the image.
3. The TKB will provide a list with all the possible relations between pairs of the selected concepts. Annotate the most representative propositions for the image.
4. Annotate VKPs: labels, arrows for parts or dynamism, colour coding/contrast, etc, and their function.

4.2 User interface

EcoLexicon data have been translated into the following Camomile collections:
1. Images are modelled as Camomile media.
2. EcoLexicon data is stored in Camomile metadata.
3. Image annotations are Camomile annotations grouped into Camomile layers.

These layers are presented to annotators in several subsequent interfaces according to our annotation guidelines to enhance consistency.

Figure 5: Home page of Manzanilla.

In Figure 5, the home page of Manzanilla is shown. After the annotator logs in, an interface appears where a search concept is entered, in this case GROYNE, and all the images related to that concept in EcoLexicon are shown (see Figure 6). The search concept is one of a list of concept entries with images already available in EcoLexicon.

The annotator then choses an image (for example Figure 7), which leads to the next interface where the image type can be chosen: photograph, drawing or flow chart (Figure 8). The use of arrows to show the direction of movement and the inclusion of procedural concepts such as LONGSHORE TRANSPORT, UPDRIFT and DOWNDRIFT, clearly show that this image is a flow chart.

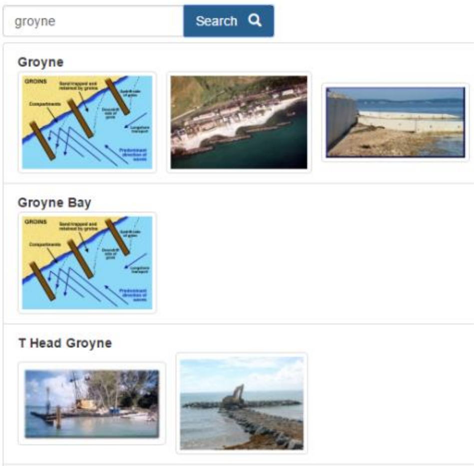

Figure 6: Concept search and image selection interface.

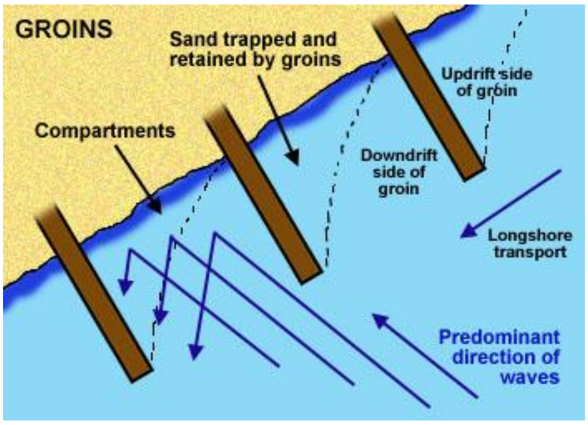

Figure 7: Image selected from EcoLexicon in search concept GROYNE.

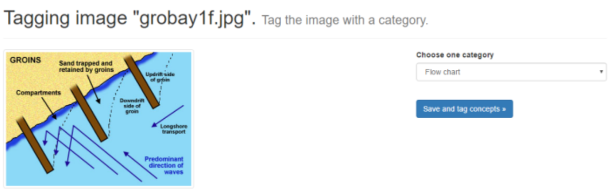

Figure 8: Image type annotation interface.

In the next interface (Figure 9), the other concepts contained in the image, apart from the initial search concept, are annotated. A list of suggestions is offered based on all the concepts that are related to the concept entry where the image is stored in EcoLexicon. New concepts can all so be tagged. These are recorded to make sure that the concept is added to EcoLexicon later on.

After tagging the concepts, in the next interface, the relations between those concepts are annotated. Suggestions are again provided based on the conceptual propositions contained in EcoLexicon. Of course, new propositions can be added as well (Figure 10).

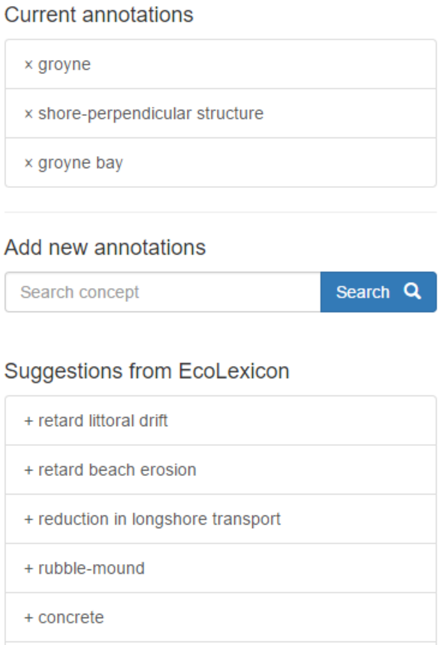

Figure 9: Extract of concept annotation interface.

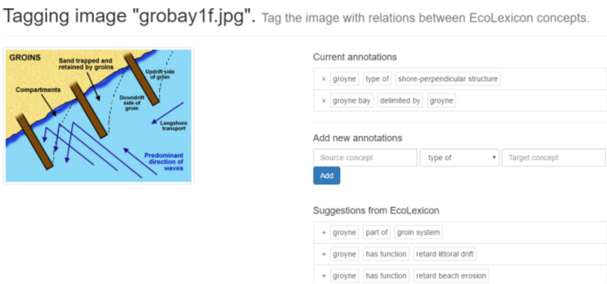

Figure 10: Conceptual relation annotation interface.

The last interface is where the VKPs used in the image are tagged (Figure 11). With the mouse, sections of the image can be marked and labelled according to type (arrow, label, colour-coding) and the function the VKP expresses in the image.

Figure 11: Interface to annotate VKPs.

At the moment, the function of each VKP is a free-text box. Thus, each annotator can freely describe the function the VKP represents in this image. However, as this will probably cause a high degree of inconsistency between annotators, a list with fixed options has been defined, which will be implemented shortly (see Table 1).

VKP	Function	
Arrow	Dynamism	Spatial
		Temporal
	Denomination	
	Delimitation	
Colour	Realism	
	Contrast	
Label	Denomination	
	Explanation	
Number	Denomination	
	Temporality	
Logical operator		
Other		

Table 1: Closed list of VKP functions.

The option “Other” has been included in case new VKPs are identified during the annotation process. In the example image of GROYNE (Figure 7), arrows are used for both dynamism, to represent the direction and movement of processes, and denomination, to show were certain entities are located (COMPARTMENTS and retained SAND). Colours are used to convey realism: the sand is beige, groynes are brown and the water is blue. Furthermore, labels are used for denomination to show where entities are located and processes take place.

5. Conclusions and future work

Manzanilla is an image annotation tool created specifically for building the visual repository of EcoLexicon. The tool will be used to annotate all existent EcoLexicon images, which will provide further insights into image description and multimodal knowledge representation.

Shortly, the tool will be evaluated to see if it provides enough annotation consistency for our purposes. A separate section will be created in the tool where 60 images will be annotated by three annotators. These annotators are all members of the research group LexiCon and are familiar with the existing literature on image description. They will be instructed on the functioning of Manzanilla during an introductory session. Then they will annotate the same images, which are selected by the authors to include a varied range of image types, concept types and uses of VKPs. Apart from inter-annotator agreement, a second more qualitative evaluation will be carried out to assess whether the image selection and annotation criteria developed are sufficient for image population of EcoLexicon or if they need more refinement.

After the evaluation and implementation of possible improvements resulting from the evaluation, Manzanilla will be made available to the public to encourage its use in other fields of knowledge.

When all existing EcoLexicon images have been annotated, the separate image repository will be developed and linked to our TKB. Then Manzanilla will be adapted to include new images. Another access route, apart from EcoLexicon concept entry, will be added to allow for direct access to newly selected images.

Although we have discarded automatic annotation for now, the Camomile framework provides active learning applications that bootstrap manual annotations and retrain or adapt the annotation system accordingly (Poignant et al. 2016, 1421). Therefore, Camomile will be able to use the dataset resulting from manual image annotation to train the system to provide semi-automatic annotation in the future.

6. Acknowledgements

This research was carried out as part of projects FF2014-52740-P, Cognitive and Neurological Bases for Terminology-enhanced Translation (CONTENT), and FFI2017-89127-P, Translation-oriented Terminology Tools for Environmental Texts (TOTEM), funded by the Spanish Ministry of Economy and Competitiveness.

7. Bibliographical References

Barriuso, A. and Torralba, A. (2012) Notes on image annotation. arXiv:1210.3448 [cs.CV].

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K. and Fei-Fei, L. (2009) ImageNet: A Large-Scale Hierarchical Image Database. *IEEE Computer Vision and Pattern Recognition (CVPR)*.

Faber, P., editor, (2012). A Cognitive Linguistics View of Terminology and Specialized Language. Berlin/Boston: De Gruyter Mouton.

Faber, P., León Araúz, P., Prieto Velasco, J. A., and Reimerink, A. (2007). Linking images and words: the description of specialized concepts. *International Journal of Lexicography*, 20(1):39–65, doi:10.1093/ijl/ecl038.

Faber, P., León Araúz, P. and Prieto Velasco, J.A. (2009) Semantic relations, dynamicity, and terminological knowledge bases. *Current Issues in Language Studies*, 1(1):1–23.

Faber, P., León-Araúz, P. and Reimerink, A. (2014). Representing environmental knowledge in EcoLexicon. In E. Bárcena, T. Read and J. Arhus, editors, Languages for Specific Purposes in the Digital Era. Educational Linguistics 19. Berlin, Heidelberg: Springer.

Jeon, J., Lavrenko, V. and Manmatha, R. (2003). Automatic image annotation and retrieval using cross-media relevance models. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 119–126, Toronto, Canada.

León-Araúz, P., Reimerink, A., Faber, P. (2013). Multidimensional and multimodal information in EcoLexicon. In A. Przepiórkowski et al., editors, Computational Linguistics (143–161). Berlin/Heidelberg: Springer.

León-Araúz, P. and Reimerink, A. (2016). Evaluation of EcoLexicon images. In F. Khan et al., editors, Joint

Second Workshop on Language and Ontology & Terminology and Knowledge Structures (LangOnto2 + TermiKS) in conjunction with the 10th edition of the Language Resources and Evaluation Conference, pages 16–22, Portorož, Slovenia.
León Araúz, P., Reimerink, A., and Faber, P. (2009). Knowledge extraction on multidimensional concepts: corpus pattern analysis (CPA) and concordances. In 8ème Conférence Internationale Terminologie et Intelligence Artificielle. Toulouse.
Mayer, R. E., Anderson, R. B. (1992). The instructive animation: helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, 84(4):715–726.
Mei, T. et al. (2008). Coherent image annotation by learning semantic distance. In IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, Piscataway, NJ, USA.
Navigli, R. & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193, pp. 217–250.
Paivio, A. (1971). Imagery and Verbal Processes. New York: Holt, Rinehart & Winston.
Paivio, A. (1986). Mental Representations: A Dual-Coding Approach. New York: Oxford University Press.
Poignant, J., Budnik, M., Bredin, H., Barras, C., Stefas, M., Bruneau, P., Adda. G., Besacier, L., Ekenel, H., Francopoulo, G., Hernando, J., Mariani. J., Morros, R., Quénot, G., Rosset, S. and Tamisier, T. (2016). The CAMOMILE Collaborative Annotation Platform for Multi-modal, Multi-lingual and Multi-media Documents. In N. Calzolari et al., editors, Proceedings of the 10th edition of the Language Resources and Evaluation Conference (LREC'16), pages 1421–1425, Portorož, Slovenia, may. European Language Resource Association (ELRA).
Prieto Velasco, J. A. (2008). Información gráfica y grados de especialidad en el discurso científico-técnico: un estudio de corpus. PhD Thesis, University of Granada.
Prieto Velasco, J. A. and Faber, P. (2012). Graphical Information. In P. Faber, editor, A Cognitive Linguistics View of Terminology and Specialized Language, pages 225–248. Berlin/Boston: De Gruyter Mouton.
Reimerink, A., León-Araúz, P. and Faber, P. (2016). Image selection and annotation for an environmental knowledge base. *Language Resources and Evaluation*, 50:1–32.
Reimerink, A., García de Quesada, M. and Montero-Martínez, S. (2010). Contextual information in terminological knowledge bases: A multimodal approach. *Journal of Pragmatics*, 42(7):1928–1950, doi:10.1016/j.pragma.2009.12.008.
Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015) ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*.
Wenyin, L., Dumais, S., Sun, Y., Zhang, H. J., Czerwinski, M. and Field, B. (2001). Semi-automatic image annotation. In Proc. of Interact 2001: Conference on Human-Computer Interaction, pages 326–333.
Xiao, T., Xu, Y.,Yang, K., Zhang, J., Peng, Y. and Zhang, Z. (2015). The application of two-level attention models in deep convolutional neural network for fine-grained image classification. arXiv:1411.6447 [cs.CV].
Zitnick, C. L. and Parikh, D. (2013). Bringing semantics into focus using visual abstraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3009–3016.

8. Language Resource References

LexiCon Research Group. (2017). EcoLexicon. Available at http/:/ecolexicon.ugr.es.
LexiCon Research Group. (2017). Manzanilla. Available at http://manila.ugr.es/apis/manzanilla/index.html.