Improving the use of electronic collocation resources by visual analytics techniques

Roberto Carlini[1], Joan Codina-Filba[1], Leo Wanner[2,1]

[1]Natural Language Processing Group, Dept. of Information and Communication Technologies
Pompeu Fabra University, C/Roc Boronat, 138, 08018 Barcelona
[2]Catalan Institute for Research and Advanced Studies (ICREA), Passeig Lluís Companys, 23, 08010 Barcelona
Email: {roberto.carlini,joan.codina,leo.wanner}@upf.edu

Abstract

With the increasing prominence of the electronic medium in lexicography, the face of collocation resources also changed. Collocation dictionaries have been extended by additional material (e.g., examples from a corpus and interfaces for targeted access to information), and tools such as Sketch Engine have been developed, which query a corpus and display the collocational (and grammatical) behaviour of a specified word. However, the paradigm of consulting, viewing and exploring the resources still follows to a major extent the traditional dictionary look up philosophy: the user enters a keyword and obtains an outcome in a sequential text format. This implies significant limitations if the user wants to contrast information concerning different keywords or their collocates, view information in incremental detail, etc. Studies on the presentation of information argue that visualization techniques facilitate comprehension. It is thus not by chance that visualization of linguistic information and data has become a popular research topic. In our work, we aim to go one step further: we research how Visual Analytics (VA), which deals with the development of techniques that support the exploration, analysis and interpretation of information, can be used to explore collocation resources in the context of learning Spanish as second language.

Keywords: collocations; active learning; visual analytics

1. Introduction

With an increasing prominence of the electronic medium in lexicography, the face of collocation resources has also changed. Collocation dictionaries have been extended by additional material such as examples from a corpus and interfaces, which allow for targeted access of information; cf., e.g., DICE `http://www.dicesp.com`. Also, tools such as Sketch Engine (Kilgarriff et al., 2014) have been developed, which query a corpus and display the collocational (and grammatical) behaviour of a specified word. However, the paradigm of consulting, viewing and exploring the resources still predomintantly follows the traditional dictionary look up philosophy: the user enters a keyword and obtains an outcome in a sequential text format. This implies significant limitations if the user wants to see which bases share a given collocate, contrast information concerning different keywords or their collocates, view in incremental detail some information, etc. Studies on the presentation of information (Tufte, 1983; Smith,

2005) argue that visualization techniques facilitate comprehension. It is thus not by chance that visualization of linguistic information and data has become a popular research topic; cf., e.g., (Collins et al., 2008, 2009; Penn and Carpendale, 2009; Feng and Lapata, 2010).

In our work, we aim to go one step further: we research how VA (Keim et al., 2008; Wong and Thomas, 2004) can be used to explore collocation resources in the context of second language learning. VA deals with the development of techniques that support the exploration, analysis and interpretation of information (in our case, collocation resources) via interactive visual interfaces.

In the context of second language learning, it is important to offer to the user the opportunity to (i) contemplate the possible collocates of a given keyword and compare the information concerning the frequency and context of their use; (ii) study the appearance of a collocation in different contexts; (iii) explore which of the keywords share the same collocate(s) and which ones do not; (iv) retrieve the syntactic structure of a collocation; etc. We explore VA techniques that account for these needs. Our resource is a large Spanish newspaper syntactic dependency corpus treebank. The corpus is indexed and processed for efficient computation of “collacability" between binary word co-occurrences, hence there holds a direct syntactic dependency and efficient access to supportive and illustrative information (such as samples of the use of a collocation in context).

In what follows, we first discuss the needs of a learner user of a collocation dictionary (Section 2). In Section 3, we then introduce the notion of VA and briefly show how it can be used for dynamic interactive exploration of collocation information. In particular, we present the VA techniques that we use in the context of the visualization of collocation information. Section 4 describes the application of these techniques to Spanish resources and illustrates their use through several examples, before Section 5 draws some conclusions from the described work and presents our future work in this area.

2. Needs of a Learner User of a Collocation Dictionary

Before we discuss the needs of the user, we shall briefly introduce the information that we assume to be available in a complete online collocation dictionary and the way it is presented.

2.1 The Content of a Collocation Dictionary

We take the Spanish collocation dictionary DiCE (Alonso et al., 2010) as an example of a complete online collocation dictionary. An entry of the DiCE contains the following main information:[1]

[1] For a complete list of the information provided in a DiCE entry, see `http://www.dicesp.com/paginas/index/1`.

1. The corresponding list of disambiguated lexemes of the lemma of the keyword (or base), together with their part of speech (PoS) and semantic category; in the case of nouns, instead of the noun tag, the grammatical gender tag is given. Consider, for illustration, the information provided for afecto 'affection':

> afecto1 m. sentimiento 'sentiment', afecto2a m. sentimiento 'sentiment', afecto2b m. manifestación 'manifestation', afecto3a adj. estado 'state', afecto3b adj. estado 'state', afecto3c adj. estado 'state'.

2. For each lexeme, as, e.g., for afecto1:

(i) its argument structure
afecto de individuo X por hecho Y 'affection of individual X for a fact Y';
(ii) its (quasi-)synonyms and antonyms
emoción, estado de ánimo, pasión1, sentimiento1a;
(iii) its subcategorization (government) structure
1 $\rightarrow$X *de* N | A_{pos}
2 $\rightarrow$Y *por* N | *ante* N | *hacia* N.

which states that the first semantic actant of afecto1 is projected onto its first syntactic actant, which is realized either as a noun with a preposition *de* 'of' or as a possessive adjective, and that its second semantic actant is projected onto its second syntactic actant, which, in turn, is realized as a noun with one of the prepositions *por* 'for', *ante* 'before', or *hacia* 'towards'.

(iv) its collocates, categorized first according to the PoS of the collocate and its default location relative to the base (i.e.,: <verb>+BASE, BASE+<verb>, <adjective>+BASE, etc.), and then, within each of these categories, according to the semantics of the collocate in combination with the base
manifestar ∼ 'manifest ∼'
expresar 'express'

The use of the individual lexemes and of the individual collocations is illustrated by examples, mainly from a corpus of Spanish of the Spanish Royal Academy (`http://corpus.rae.es/creanet.html`); consider Figure 1 for illustration.

2.2 The Needs of the User

Online collocation dictionaries of the type of DiCE facilitate information when the intention of the user is to look up the collocates of a base (in order to then choose one of them), to verify a collocation they had in mind, or to learn about the use of a specific collocation in context. They may also provide some detailed information on the base lexeme—e.g., its argument and subcategorization structures or its (quasi-)synonyms or antonyms. To obtain

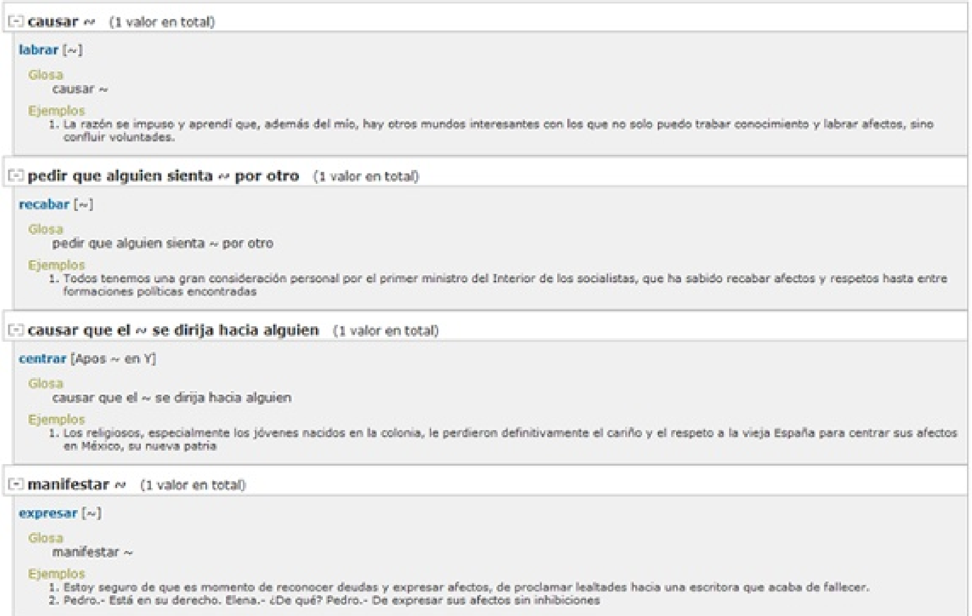

Figure 1: Display of the verb+noun collocations of afecto1 'affection' in DiCE

the desired information (in a sequential text format), the user needs either to introduce the base into an interface or select it (possibly in a cascaded menu) from a list (as is the case in DiCE). However, this traditional dictionary look up philosophy is not sufficient when the user is a language learner and the dictionary is supposed to serve as an instrument that supports active learning. Active learning is closely related to exploration and even more so in the context of active learning of collocations: collocations are idiosyncratic in that two bases with similar meanings may have different collocates (possibly with the same semantics; cf., e.g., *labrar afecto* 'produce affect' vs. *inspirar simpatía* 'inspire sympathy') or share the same collocates (as, e.g., *té* 'tea' and *café* 'coffee': *tomar un té/café*), deviate from a literal translation from L1 (as, e.g., *take* [*a*] *walk* vs. *dar* [*un*] *paseo*, lit. 'give a walk') or not (as, e.g., *give* [*a*] *talk* vs. *dar* [*una*] *conferencia*), etc. This can only be learned by navigating in the collocation spaces, by comparing, clustering, etc.

The most intuitive questions to explore in view of a collotation include, for instance:[2]

- Which other lexemes collocate with the base of this collocation and how common are these collocations (either compared to the given collocation or in absolute terms)?
- Which other bases take the collocate of this collocation (and, again, how common are these collocations)?
- What is the overlap of collocates of the given base with semantically similar bases?
- What is the typical context of this collocation?

[2] These and further similar questions can be derived from the didactic studies related to collocation learning; see, among others, (Hausmann, 1984; Lewis, 2000; Higueras García, 2011).

We shall now investigate how VA can help to explore these or similar questions and to provide the information that the user expects to encounter when consulting a collocation dictionary such as DiCE.

3. Visual Analytics Techniques and Collocation Information

In what follows, we first give a short introduction to Visual Analytics and then discuss techniques that we consider appropriate for the display and exploration of collocation information.

3.1 What is Visual Analytics?

Visual Analytics (VA) is a recent research area that emerged within the field of information visualization as a response to the need of (possibly unexperienced) users to explore new (usually large) information spaces; cf.: "Visual analytics is the formation of abstract visual metaphors in combination with a human information discourse (interaction) that enables detection of the expected and discovery of the unexpected within massive, dynamically changing information spaces." (Wong and Thomas, 2004). Indeed, this is exactly what is expected by a learner who actively explores the "collocation space". A great number of different visual metaphors have been proposed by the VA community for the exploration of different types of information spaces; see, e.g., `http://d3js.org/` for an extensive library. Among the most common visual metaphors are various types of networks (to visualize the connectivity between the elements of the explored space), trees (to visualize hierarchical relations between the elements of the space), flows (to visualize the change of the information space over a time line), glyphs (to visualize multidimensional data), etc. Figure 2 presents a fragment of a radial tree taken from `http://bl.ocks.org/mbostock/4063550`, cited in (Butt and Culy, 2014).[3] The tree is interactive in that it can be collapsed, expanded, zoomed-in, etc.

The general principle underlying nearly all metaphors is "Overview first, zoom and filter, then details-on-demand" (Shneiderman, 1996). To facilitate an overview, data tend to be aggregated (clustered) with respect to specific features. The zoom allows for inspection of specific patterns or subsets of data by applying filtering. The "details-on-demand" displays the individual features, examples, etc. related to individual entities in the information space.

We shall now discuss how VA can be used for visualizating and exploring collocation information.

[3] Note that Figure 2 does not represent any collocation-oriented information; it is displayed just for the sake of the illustration of the notion of a radial tree.

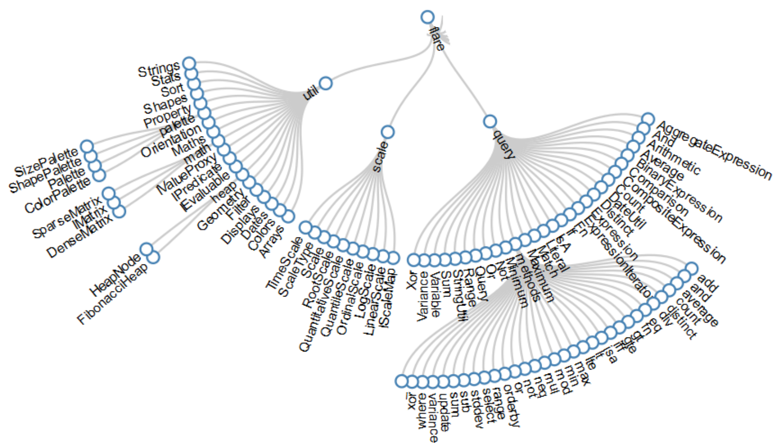

Figure 2: Example of a fragment of a radial tree visualization

3.2 Visualization Analytics and Collocation Information

One could imagine using radial trees as shown in Figure 2 for the visualization of collocation information. But the most appropriate visual metaphor of collocations in context is a network or graph. Firstly, a base combines with several collocates, while several bases as a rule share one or several collocates. This results in a connected structure with two types of nodes, bases and collocates. Secondly, the frequency of the co-occurrence of a base with a collocate in a corpus, which indicates how common a collocation is, can be expressed by the design of the arc between the base and the collocate or the size of the collocate node. Furthermore, to express, for instance, that some bases share certain collocates or some collocates co-occur with several bases, the nodes in the network can be visually clustered into hypernodes.

For casting collocation information into a network, we draw upon techniques used for community detection in social networks. In particular, we use *Gephi* (Bastian et al., 2009), an off-the-shelf network design workbench. *Gephi* is a software for network visualization and analysis written in Java and is thought to “help data analysts to intuitively reveal patterns and trends, highlight outliers and tell stories with their data”.[4] It combines a powerful set of built-in capabilities to explore, analyze, spatialize, filter, cluster, manipulate and export all types of graphs, and is provided with an open API that allows users and developers to write their own plug-ins in order to extend the software.

Gephi software can be used through a GUI as an interactive program since it follows the “visualize-and-manipulate” paradigm and was designed specifically for VA, i.e., for exploration of data. However, a meaningful interaction directly with *Gephi* requires knowledge regarding the explored data, basic notions of network design, the available transformations and their effect, the visualization layouts and how they can be tuned, etc. In short, it is

[4] *Gephi* is free and distributed under the GPL 3.

intended for specialists (e.g., data analysts), not (potentially formally untrained) end users such as language learners.

Therefore, to ensure an "easy-to-follow" dynamic interaction, we use *Gephi* as a library. We first generate static visualizations of graphs, export the resulting *Gephi* graph in gefx format, and subsequently visualize it by means of a web interface that we developed using the `sigma.js` JavaScript library[5]. `sigma.js` is among the best graph drawing JavaScript libraries available. Besides being easily customizable and having a lot of built-in features, such as Canvas and WebGL renderers or mouse and touch support, it provides a plugin system, so anybody can add code to implement any other functionality.

4. Towards Visual Exploration of Collocation Spaces

In order to be able to offer the functionality of the exploration of collocation resources as sketched in Subsection 2.2 above, these resources need to be preprocessed in several terms. Therefore, before we embark on the description of the implementation of VA, we present the preprocessing of the resources we use.

4.1 Preprocessing of the Collocation Resource

Our resource is a large Spanish newspaper syntactic dependency corpus treebank. The treebank has been indexed in Solr, with each sentence being captured in the index in three different ways in order to be able to retrieve the following kinds of information:

1. The sentence as it appears in the original corpus.
2. The sentence as a sequence of "PoS|lemma" tags, to allow for searches based on the lemma with a given PoS.
3. Sequences of lemmas with their parents in the dependency tree: For each lemma in the sentence, an element that includes the term, its PoS, the PoS of the parent, the lemma of the parent, the syntactic relation between the lemma and the parent and the position of both in the sentence. When there is a preposition between a verb and a noun, the preposition is removed and a direct relationship is created. This structure allows for searches such as a "lemma being a noun related to any verb" and finds these verbs and how they are related.

For the treebank's binary word co-occurrences between which a direct syntactic dependency holds, the *collocate-weighted normalized pointwise mutual information* (NPMI_c) has been calculated as a measure of "collocalibity"; cf. (Carlini et al., 2014).[6] Solr's faceted search has been

[5] `http://sigmajs.org/`.

[6] In contrast to the standard PMI, as commonly used in Corpus Lexicography since (Church and Hanks, 1989), NPMI_c takes the asymmetric nature of collocations into account.

used in order to retrieve the information needed for the computation of the NPMI$_c$s, which are precomputed and stored in a relational database. The use of the relational database and Solr facilitates efficient access of individual tokens, lemmas, token/lemma – co-occurrences with NPMI$_c$s, syntactic dependencies, and example sentences (with their dependency structures) and real-time delivery of the corresponding information (including examples) via the user interface.

4.2 Realizing VA for Collocation Resource Exploration

In Subsection 2.2, we listed some questions concerning both individual collocations and collocation collections the exploration of which should be facilitated by use of a VA tool. In what follows, we present some of our realizations aimed to fulfil this demand.

4.2.1. Exploring the collocation space of a base

In order to help the learner to explore the collocability of a given base, the collocates of this base are clustered with respect to their context (and thus with respect to their distributional semantics) and displayed in terms of coloured circles. The size of a collocate's circle indicates the commonality of the collocation formed by the base–collocate co-occurrence (more precisely, its size is proportional to its NPMI$_c$). Each cluster is displayed in a different colour. Figure 3 illustrates this kind of visualization for the collocation space of *té* 'tea'. *Beber* 'drink' (cf. *beber té* 'drink tea') and *tomar* 'take' (cf. *tomar té*, lit. 'take tea') form one cluster (as a matter of fact, *beber* and *tomar* are synonymous in their role as collocates of *té*). A second, considerably more heterogeneous, cluster is formed by *preparar* 'prepare', *ofrecer* 'offer', *pedir* 'ask for', *servir* 'serve', and *compartir* 'share'.

Café 'coffee' can be expected to share as base its collocates with *té*. However, given that, on the one hand, drinking coffee in Spain is much more common than drinking tea and, on the other hand, *café* is polysemous in that it can also refer, e.g., to a location or to a drink after lunch in general, the graph for *café* is considerably richer; cf. Figure 4. Thus, it also contains clusters related to breakfast (*desayunar* 'have breakfast'), to the social event of drinking coffee (*invitar* 'invite', *compartir* 'share'), which overlaps with the cluster of *café* as location (*frequentar* 'frequent'), and to coffee as a plant (*plantar* 'plant'), etc.

To obtain a graph such as that of *té* or *café*, we first generate a weighted graph of nodes centred on the base, with all of its collocates that show a NPMI$_c$ over a given threshold.[7] The weighted graph is then clustered using the modularity algorithm presented in (Blondel et al., 2008) and as implemented by (Lambiotte et al., 2008) in *Gephi*.

[7] We set the threshold to 0.2 since even if an NPMIc higher than 0 indicates that the relation between both elements is beyond randomness, more significance is needed for the two elements to become a collocation and avoid noise.

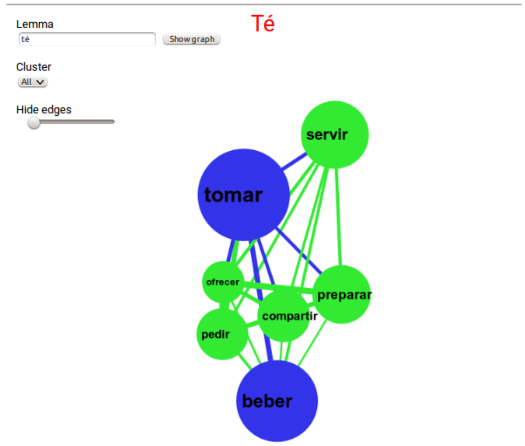

Figure 3: Collocation space of the base *té* 'tea'

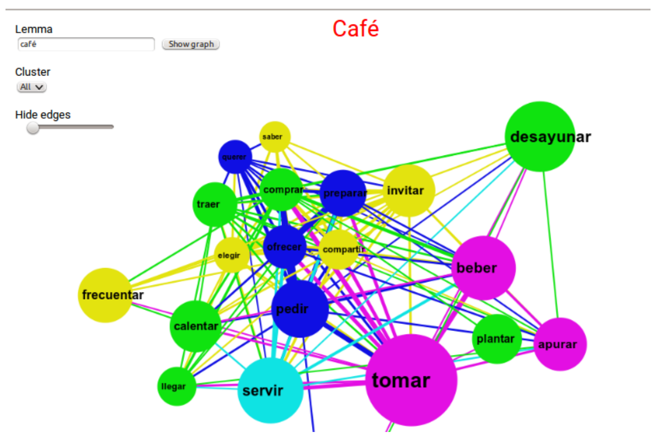

Figure 4: Collocation space of the base *café* 'coffee'

4.2.2. Collocation space of bases sharing collocates

In order to move on from the exploration of the collocation space of a single base to a (contrastive) exploration of the space of several bases in parallel, the weighted graph from above is expanded by all bases of the collocates that are related to them with a NPMI_c above the threshold via a specific syntactic dependency relation (e.g., direct object).[8] With this action we obtain a bipartite graph of bases and collocates.

In a second step, we use *Gephi*'s multimodal transformation to find for every pair of collocates how many bases they have in common.[9]. This produces a reduced graph where only the collocates are present. However, as a rule, it is still a high density graph that is difficult to

[8] In the current initial version of our VA experiments, we work by default with some prominent dependency relations such as 'direct object', 'indirect object' and 'subject'. It is foreseen that the learner can choose the relations interactively via the interface.

[9] https://marketplace.gephi.org/plugin/multimode-networks-transformations-2/

view and inspect. Therefore, the edges under a certain threshold are pruned to simplify the graph.[10] For the spatial distribution of nodes, a force atlas is used and labels are adjusted to avoid label superposition. Once the collocates are clustered, the graph is expanded again with the bases.

Finally the elements in the graph are scaled such that:

- the size of the bases is the sum of the NPMI$_c$s they have with the different collocates with the NPMI$_c$ of the collocate; in this way, bases that highly correlate with the source base appear bigger;
- the strength of the edges between collocates indicate how many bases they have in common;
- the strength of the edges between bases and collocates is proportional to their NPMI$_c$s.

In Figure 5, the collocate selected to be in focus (*beber* 'drink') is represented as a hexagon, the other collocates as circles and the bases as triangles. The bigger the size of a base triangle the more collocates it shares with *té*.

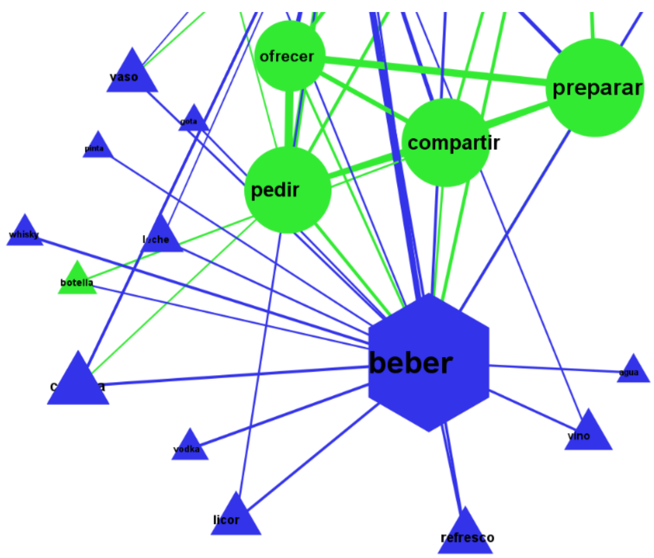

Figure 5: Collocation space of several bases related with *beber* and the original base *té*

4.2.3. Zooming in on the details of a collocate or collocation

The user may also want to further explore individual elements of the graph. This can be done using the "zoom-in" function. Thus, the user can, e.g., click on a collocate and obtain

[10] After a series of tests, we set this threshold to 0.3, as it sufficiently reduces the density of the data.

information about it, get sample sentences with the use of the collocation formed by the collocate and the corresponding base, and the information regarding which other bases this collocate co-occurs (as in the initial setting, only those bases are displayed that have an NPMI_c above the threshold). For instance, if we click on *apurar* '[to] drain' we learn (see Figure 6) that *apurar* co-occurs with such bases as *cerveza* 'beer', *vaso* 'cup', and *copa* 'glas'. Several examples from the corpus illustrate the use of *apurar* in context (in this case, its co-occurrence with *café*). Also, the learner can learn about the frequency of the collocate in the corpus and its NPMI_c.

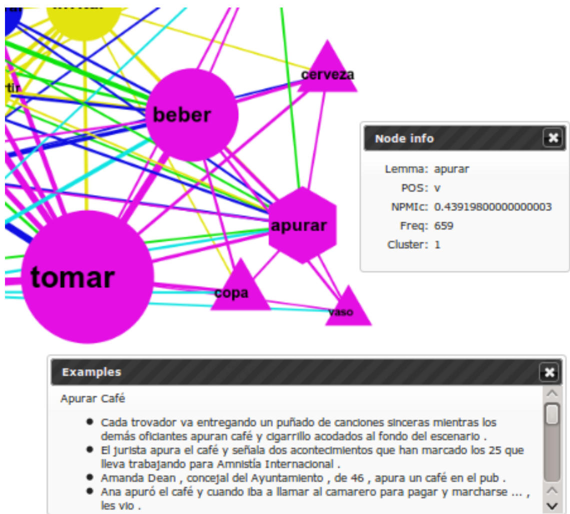

Figure 6: Zooming in on the collocate 'apurar' '[to] drain'

The user can also zoom-in on the link between a base and a collocate (i.e., on a specific collocation) to obtain examples; see Figure 7, where the user clicked on the link between *vaso* 'cup' and *apurar* 'drain' to obtain sample sentences in which *vaso* and *apurar* appear as collocation. The information regarding which collocates belong to the same cluster as *apurar* (namely *tomar* 'take' and *beber* 'drink') and with which prominence, and which other prominent clusters are involved in the collocation space of *vaso* (in this case, the cluster consisting of *servir* 'serve'), is also displayed.

4.2.4. Navigation within collocation spaces

The user can navigate starting from the graph centred around a given base to a graph centred around one of the bases with which this base shares some of the collocates. This is done by double-clicking on the base the user wishes to look at. The obtained graph is obviously different from the starting graph because it is centered around the new base. Figure 8 shows the graph for *copa* 'cup', obtained departing from the graph of *café* 'coffee'. The most prominent collocates for *copa* remain (as already for *café*) *beber* 'drink' and *tomar*

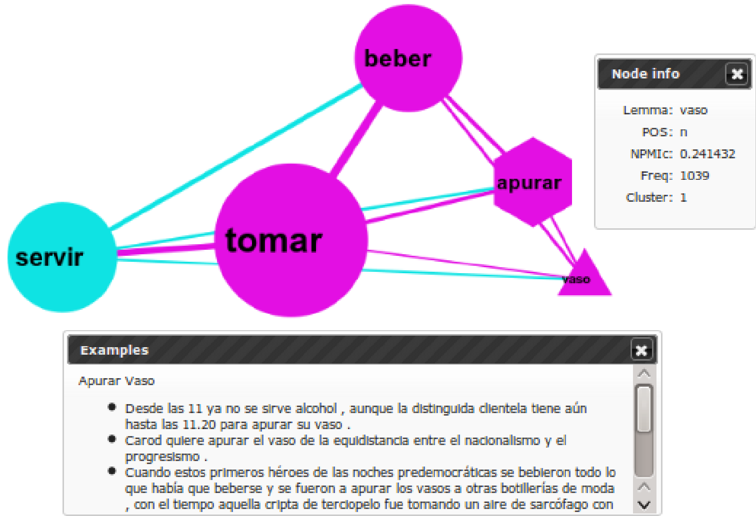

Figure 7: Zooming in on *apurar* [*el*] *vaso* 'drain [the] cup'

'take', but it can also be observed that *copa* has a number of collocates not shared with *café*. In this context, it is important to notice that in all of the given graphs, the similarities and correspondences between bases are always calculated and displayed with respect to the subset of the collocates of the base in focus, not with respect to the full language model.

4.2.5. Exploration of collocate clusters

In Subsection 4.2.1, we already mentioned that collocates are clustered in accordance with their distributional semantics. An ideal clustering algorithm would group collocates with respect to a theoretically well-defined semantic collocate typology—as, e.g., *the lexical functions* (LFs) (Mel'čuk, 1996) or a generalization thereof. In DiCE, the glosses of the collocate groups in the individual entries for the bases (see Figure 1) are, in fact, LFs.[11] For automatic classification of given collocation lists in terms of LFs, see, e.g., (Wanner, 2004; Wanner et al., 2006; Moreno et al., 2013). In the current implementation of our VA tool, collocates are clustered according to the strength of the relationships between them (number of common bases) using the "Louvain algorithm" (Blondel et al., 2008) for community detection. This algorithm is graph-based and tries to optimize the modularity of the community.[12] Applied to the collocates, it groups those collocates that share more bases between them than with the other collocates.

Each base is assigned to the cluster of the collocates which show, in the co-occurrence with it, a NPMI_c higher than the threshold. The user can restrict the visualization of the graph

[11] The interface of the DiCE also allows for the display of actual LF labels, along with the glosses.
[12] Modularity measures the relation between the density of edges inside communities to edges outside communities.

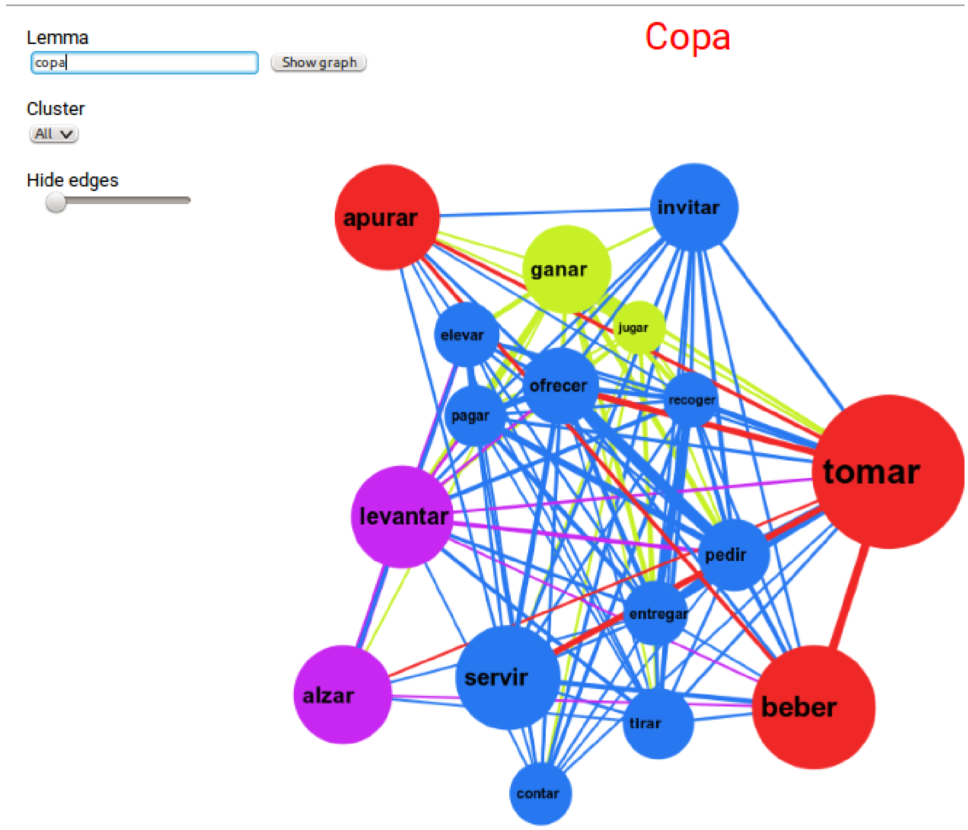

Figure 8: Navigating from *café* to *copa* 'cup'

to a subset of nodes belonging to a single cluster. Figure 9 shows the resulting graph after selecting one of the clusters from the graph for *copa* 'cup'.

5. Conclusions and Future Work

In this paper we presented some VA techniques for dynamic interactive exploration of collocation information, starting from the collocation space of a single base and either expanding it to the space of several bases or zooming in on the details of a single collocation. We believe that VA is crucial in all active learning environments, but particularly so in a collocation learning environment since collocations are idiosyncratic in their nature and thus require extra support for memorization.

The interface of the current implementation of our VA tool has been first realized as a standalone web application. It is now about to be built into the HARENES project interface (Wanner et al., 2013), where it will be integrated with other functionalities—for instance, that the learner can introduce a collocation, validate its correctness and obtain correction suggestions in case it is not correct, or introduce a whole text and receive correction sugges-

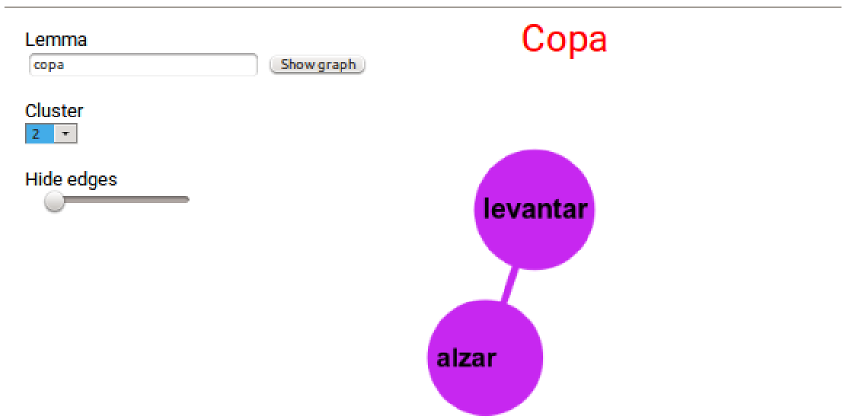

Figure 9: Selection of the collocates that belong to a cluster composed of the collocates *levantar* 'raise' and *alzar* 'lift' (in connection with *copa* 'cup' as a glass or as a trophy)

tions for the detected miscollocations. In this context, we plan also to experiment with the use of other collocate clustering (or classification) algorithms than the one that is used in the current VA tool—for instance, the one described in (Moreno et al., 2013).

The presented tool can be built into the interface of any online collocation dictionary such as DiCE, where it could be used to better visualize and explore the information that is available in this dictionary. However, prior to this integration, it must be evaluated—ideally in real language learning settings, involving students and teachers. Furthermore, it should be kept in mind that its current design does not necessarily follow rigorous didactic and/or visualization optimization considerations. A collaboration of specialists from these fields will be necessary to make the presented prototypical implementation a valuable aid in second language collocation learning.

6. Acknowledgements

The work presented in this paper has been supported by the Spanish Ministry of Economy and Competitiveness under the contract number FFI2011-30219-C02-02 in the framework of the HARENES Project, carried out in collaboration with the DiCE Group of the University of La Coruña.

7. References

Alonso, M., Nishikawa, A., & Vincze, O. (2010). DiCE in the web: An online Spanish collocation dictionary. In Granger, S. & Paquot, M., editors, *Proceedings of eLex 2009, Cahiers du Cental 7*. Louvain-la-Neuve. Presses universitaires de Louvain, pp. 367–368.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. In *Proceedings of the International AAAI Conference on Weblogs and Social Media*. Menlo Park. AAAI Press.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*.

Butt, M. & Culy, C. (2014). Visual analytics for linguists. ESSLI '14 Course material `http://www.sfs.uni-tuebingen.de/~cculy/courses/ESSLLI2014/`.

Carlini, R., Codina-Filba, J., & Wanner, L. (2014). Improving collocation correction by ranking suggestions using linguistic knowledge. *NEALT Proceedings Series Vol. 22*.

Church, K. & Hanks, P. (1989). Word Association Norms, Mutual Information, and Lexicography. In *Proceedings of the 27th Annual Meeting of the ACL*, pp. 76–83.

Collins, C., Carpendale, S., & Penn, G. (2009). DocuBurst: Visualizing Document Content Using Language Structure. In *Proceedings of the Eurographics/IEEE-VGTC Symposium on Visualization (EuroVis '09)*. Eurographics Association, pp. 1039–1046.

Collins, C., Penn, G., & Carpendale, S. (2008). Interactive visualization for computational linguistics. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies*, Morristown, NJ, USA. Association for Computational Linguistics.

Feng, Y. & Lapata, M. (2010). Visual Information in Semantic Representation. In *Proceedings of the 2010 Annual Conference of the North American Chapter of the ACL and the Conference on Human Language Technologies*. Morristown, NJ, USA. Association for Computational Linguistics, pp. 91–99.

Hausmann, F.-J. (1984). Wortschatzlernen ist kollokationslernen. zum lehren und lernen französischer wortwendungen. *Praxis des neusprachlichen Unterrichts*, 31(1), pp. 395–406.

Higueras García, M. (2011). Lexical collocations and the learning of Spanish as a foreign language. In Cifuentes Honrubia, J. and Rodríguez Rosique, S., editors, *Spanish Word Formation and Lexical Creation*. Benjamins Academic Publishers, Amsterdam, pp. 439–464.

Keim, D., Mansmann, F., Schneidewind, J., Thomas, J., & Ziegler, H. (2008). Visual Analytics: Scope and Challenges. In Simoff, S., editor, *Visual Data Mining*, LNCS 4404. Springer Verlag, Berlin, pp. 76–90.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). *The Sketch Engine: ten years on.* Lexicography: Journal of ASIALEX, 1(1), pp. 7–36.

Lambiotte, R., Delvenne, J.-C., & Barahona, M. (2008). Laplacian dynamics and multiscale modular structure in networks. *arXiv preprint arXiv:0812.1770*.

Lewis, M. (2000). *Teaching Collocation. Further Developments in the Lexical Approach*. LTP, London.

Mel'čuk, I. (1996). Lexical functions: A tool for the description of lexical relations in the lexicon. In Wanner, L., editor, *Lexical Functions in Lexicography and Natural Language Processing*. Benjamins Academic Publishers, Amsterdam, pp. 37–102.

Moreno, P., Ferraro, G., & Wanner, L. (2013). Can we determine the semantics of collocations without using semantics? In *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of eLex 2013: Electronic Lexicography in the 21st Century*. Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, pp. 106–121.

Penn, G. & Carpendale, S. (2009). Linguistic information visualization. ESSLI '09 Course material. `http://esslli2009.labri.fr/documents/carpendale_penn.pdf`.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*, pp. 336–343.

Smith, K. L. (2005). *Handbook of visual communication: theory, methods, and media*. Rutledge, Oxford.

Tufte, E. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, CN, USA.

Wanner, L. (2004). Towards automatic fine-grained semantic classification of verb-noun collocations. *Natural Language Engineering Journal*, 10(2), pp. 95–143.

Wanner, L., Bohnet, B., & Giereth, M. (2006). Making Sense of Collocations. *Computer Speech and Language*, 20(4), pp. 609–624.

Wanner, L., Verlinde, S., & Alonso Ramos, M. (2013). Writing assistants and automatic lexical error correction: word combinatorics. In *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of eLex 2013: Electronic Lexicography in the 21st Century*. Trojina, Institute for Applied Slovene Studies/Eesti Keele Instituut, pp. 472–487.

Wong, P. C. & Thomas, J. (2004). Visual analytics. *IEEE Computer Graphics and Applications*, 24(5), pp. 20–21.