

Testing an electronic collocation dictionary interface: *Diccionario de Colocaciones del Español*

Orsolya Vincze, Margarita Alonso Ramos

Universidade da Coruña, Campus da Zapateira s/n, A Coruña 15071, Spain
E-mail: ovincze@udc.es, lalonso@udc.es

Abstract

This paper describes the results of a usability study that tests the online interface of the *Diccionario de Colocaciones del Español* (DiCE). This dictionary was conceived with the purpose of providing a detailed description of Spanish collocations in accordance with the theoretical guidelines of the Explanatory and Combinatorial Lexicology. Although from the outset of dictionary compilation, accessibility of the DiCE interface has always been taken into account, no usability test has been carried out to see how different target user groups are able to perform with the dictionary. Our aim was to assess the functionality of the different search options offered by the interface, both in terms of their efficiency and the adequacy of presentation from the point of view of the user. As the results of the test show, the overall quality of interaction between users and the dictionary was good, although we have also identified some areas for improvement, which are provided as design recommendations in the concluding part of the paper.

Keywords: usability testing; collocation dictionary; explanatory and combinatorial lexicology, search strategies, log file analysis

1. Introduction

The present paper describes the results of a usability study that tests the online interface of the *Diccionario de Colocaciones del Español* (DiCE, Alonso Ramos, 2004, Alonso Ramos et al., 2010 and Vincze et al., 2011). This dictionary was conceived with the purpose of providing a theoretically well-founded and detailed description of Spanish collocations. However, it was always intended as a useful tool for its users. It is for this reason that, after different modifications of the interface, we decided to carry out a usability test to see how different target user groups are able to perform with the dictionary. Our aim was to assess the different search options offered by the interface both in terms of their efficiency and the adequacy of their presentation from the point of view of the user. In the next section we briefly overview similar dictionary usability studies. Subsequently, we present our own study and the conclusions drawn from the results obtained.

2. Dictionary usability studies

Various aspects of dictionary use are studied. Most studies aim to decipher for which purposes dictionaries are used, what knowledge or abilities dictionary users have or require, or how dictionaries contribute to language learning. Heid (2011) proposes a different approach: the application to electronic dictionaries of *usability testing*, as defined by information science. This line of research implies testing dictionaries at

the level of functionality, much like in the case of other kinds of software tools. Studies that have applied usability testing methodology include Heid and Zimmerman (2012), which compares different types of access to collocations in mock-up dictionary interfaces, and Hamel (2012), which provides a detailed description of a usability experiment with a dictionary prototype concentrating on lexical selection, combination and paraphrase. Jousse et al. (2011) reports briefly on a test performed on a prototype collocation dictionary developed following the same theoretical framework as DiCE (see below), without providing quantitative results.

3. The study

3.1 The interface tested

The *Diccionario de Colocaciones del Español* (DiCE) is an online collocation dictionary of Spanish, which has been designed in accordance with the postulates of the Explanatory Combinatorial Lexicography (Mel'čuk et al., 1995), and is mainly oriented to language production. The DiCE represents collocations as restricted combinations of two lexical elements: the *base*, the element with more semantic weight which is freely selected in language production, and the *collocate*, an element whose selection is conditioned by lexical restrictions imposed by the base. For instance, in the combination *reanudar una amistad* 'renew a friendship', the noun is the base, and it conditions the selection of the collocate verb.

In order to offer dynamic access to the information stored in the DiCE database, in addition to the *dictionary module*, the current user interface incorporates various advanced search options. Each of these was conceived to provide the user with a more direct path of access to a specific type of information. Since the main objective of the usability test was to measure the functionality of the different search options, we provide a brief description of these.

1. *Dictionary module*: This option provides a traditional collocation dictionary type access to combinatorial information. The entry of each lemma contains the subentries of its corresponding lexical units, where collocations are grouped according to their syntactic pattern and semantic content.
2. *Advanced search module*:
 - a. *What does it mean?*: This reception-oriented module provides direct access to the entry of a specific collocation. The user is prompted to introduce a base (e.g. *amistad*) and a collocate (e.g. *reanudar*).
 - b. *Writing aid*: This is a production-oriented module, which allows the user to find collocates of a given base (e.g. *amor* 'love'), corresponding to a specific part of speech, and a meaning (e.g. 'felt for one another'), such as *amor mutuo* 'mutual love'.
 - c. *Direct search*: This option allows finding collocations in DiCE encoded

by a specific Lexical Function (Mel'čuk et al., 1995) (e.g. Sing(remordimiento) = *acceso de* ~ 'fit of remorse').

- d. *Inverse search*: This last module prompts the user to introduce a collocate (e.g. *cumplir* 'fulfill') in order to find the bases with which it can be combined (e.g. *deseo* 'wish', *esperanza* 'expectation').

3.2 The questionnaire

The questionnaire used in the usability test consisted of 13 questions. Participants were instructed to conduct searches on the dictionary interface in order to retrieve the answer for each item, even if they did feel able to provide a solution relying only on their own knowledge. Questionnaire items were designed in such a way that, although in most cases they could be resolved via navigating the dictionary module, the most direct path to obtain an answer was through using the advanced search options. In Figure 1 we show a few sample questions together with the optimal query type to be used. The numbers in brackets indicate the number of questionnaire items corresponding the given query type; note that items indicated as optimally searched by the same query type are not formulated in exactly the same way. Following the usability test itself, a brief post-test questionnaire was administered in order to measure user satisfaction.

<p>What verbs can be used with the lexical unit <i>cariño</i> 2 'affection'?</p> <ul style="list-style-type: none"> ○ optimal query type: Dictionary module/Writing aid (2) <p>What does <i>reanudar la amistad</i> 'renew a friendship' mean?</p> <ul style="list-style-type: none"> ○ optimal query type: What does it mean? (4) <p>Find the adjectives you can use to speak about <i>amor</i> 'love' 'that is felt for one another'</p> <ul style="list-style-type: none"> ○ optimal query type: Writing aid (3) <p>Find the collocates of <i>remordimiento</i> 'remorse' codified by the Lexical Function <i>Sing</i>.</p> <ul style="list-style-type: none"> ○ optimal query type: Direct search (2) <p>Find all collocations with the verbal collocate <i>cumplir</i> 'fulfill'.</p> <ul style="list-style-type: none"> ○ optimal query type: Inverse search (2)

Figure 1: Sample questionnaire items

3.3 Participants

The 26 informants who participated in the study represent four groups of different target user-profiles of DiCE: 1) Eight informants are Spanish university students. They represent a group of native Spanish users with certain language awareness. 2) Nine participants are foreign university students majoring in Spanish. These informants are upper-intermediate or advanced learners of Spanish as L2. 3) Five

informants are teachers of Spanish or English as a foreign language, all of them native speakers of Spanish. 4) Finally, the last five informants are Spanish PhD students of translation studies, all native speakers of Spanish. As a group, they can be considered as language professionals, characterized by an elevated language awareness and considerable expertise in the use of lexicographic tools.

3.4 Procedure

The experiment can be divided into three main phases: an informative session, the usability test proper, and a post-test questionnaire. Previous to completion of the usability questionnaire, the participants received a brief introduction to the concept of collocations, and were given some instructions on the completion of the usability test; however they were not instructed in the use of DiCE. After having received all necessary information, participants completed the usability questionnaire on their home computers. They were asked to provide the IP address of their computer, and the time and date of connection, so that their actions could be tracked in the DiCE website log files.

3.5 Data analysis

For quantitative analysis of the results of the usability test, we adopted the criteria described in e.g. Nielsen (1993). The usability of an interface can be measured along three main aspects: *effectiveness*, *efficiency* and *user satisfaction*. *Effectiveness* of the interaction can be measured through the task outcome, in our case, the participants' performance on the usability questionnaire represented by the number of correct answers provided.

Efficiency of the interaction is measured through task duration and the efforts of the user to accomplish the task, i.e. the degree of interaction with the dictionary interface. In our case, we established three parameters for measuring efficiency: 1) the *net time* required to complete the query in the case of each individual test item; 2) the *effort measure* calculated as the sum of the number of times a specific search option is chosen by the participant, the number of times a search filter is set, and the number of times the participant hits the *Search* button before obtaining the definitive answer for the test item; and 3) *query-type adequacy* based on the search option used to retrieve a correct answer. Here, 3 points were assigned when the participant used the most optimal search option for the question with all filters correctly set; 2 points when they used one of the advanced search options – though not the most adequate one – or when they failed to optimally set some of the search filters; and 1 point when they used the dictionary module in place of another search option which would have provided a more direct access to the information.

While effectiveness and efficiency constitute objective measures, and can be assessed on the basis of participants' answers to the items of the usability questionnaire together with the data obtained from the log files, the third aspect, *user satisfaction*,

being a subjective indicator, is evaluated on the basis of the results of the post-test questionnaire.

4. Results

The mean number of correct answers provided per participant was 9.62, with a standard deviation (SD) of 3.35, out of the total number of 13 questions. Four participants out of the 26 succeeded in finding the correct answer for all questions and 10 participants answered 11 or 12 questions correctly. Two participants only provided one correct answer, before deciding not to continue with the test.

From the efficiency scores (see Table 1), we can conclude that participants who obtained 12 or more correct answers tended to need less time, made less effort per query and simultaneously used the more adequate access path more often than others; a fact that suggests that they can be considered more skillful users. Note that both mean time and effort indicate the difficulty faced as a result of the participants' unfamiliarity with the user interface. Another tendency that can be observed in the data is that often participants who obtained a higher query adequacy score made more effort. This may be a result of these participants tending to experiment more with the different search options available on the DiCE interface, and managing to find the more straightforward ways to access information. Indeed, these participants provided more correct answers than users who tended to employ almost exclusively the more basic traditional dictionary-type access.

	Net time	Net time per test item	Total efforts	Efforts per test item	Query-type adequacy
1-4 corr. ans. (n=3)	28:39	03:39	196.67	15.13	2.33
SD	18:32	00:54	167.41	12.88	1.15
7-9 corr. ans. (n=6)	44:58	03:28	264.83	20.37	1.87
SD	20:14	01:33	107.68	8.28	0.71
10-11 corr. ans. (n=9)	50:18	03:53	355.67	27.36	2.36
SD	28:06	02:10	142.31	10.95	0.52
12-13 corr. ans. (n=8)	25:49	02:06	202.63	15.59	2.60
SD	11:57	00:59	56.61	4.35	0.48
MEAN	39:02	03:12	269.27	20.71	2.32
SD	22:54	01:42	129.17	9.74	0.66

Table 1: Summary of overall task efficiency

The number of participants who managed to find the correct answer, together with efficiency measures for each group of questions representing a specific anticipated optimal query type, provides information on which items of the usability questionnaire were especially problematic (see Table 2).

	Correct answers	Net time	Efforts	Query-type adequacy
Dictionary/ Writing aid (Qs 1, 10)	15.50 SD=0.71	03:26 SD=03:06	17.58 SD=16.84	2.94 SD=0.25
What does it mean? (Qs 2, 4, 11, 13)	23.33 SD=1.53	02:23 SD=03:05	18.15 SD=22.68	2.24 SD=0.96
Writing aid Qs 3, 6, 12)	17.00 SD=5.29	03:10 SD=3:00	20.07 SD=18.96	1.82 SD=0.99
Direct search (Qs 7, 9)	19.50 SD=2.12	03:35 SD=04:14	26.29 SD=22.72	2.23 SD=0.84
Inverse search (Qs 5, 8)	18.50 SD=0.71	04:27 SD=04:22	31.27 SD=25.90	2.78 SD=0.48

Table 2: Summary of effectiveness and efficiency for question groups according to optimal query type

Participants were most successful in answering questionnaire items which were categorized as most suitable for the *What does it mean?* search option. They also needed the least time, and made on average less effort than in the case of most other questionnaire items. The second highest mean of correct answers was achieved in the case of items which were classified as optimally queried using *Direct search*, despite the fact that these involved the use of Lexical Functions, with which participants were not familiar. A slightly lower number of participants answered correctly in the case of the questions which prompted finding collocations in the dictionary starting from the collocate, and could be resolved using the *Inverse search* option. Note that these were the only questionnaire items where participants necessarily had to make use of a specific advanced search option; whereas the answers to all other items could be queried using the *Dictionary module*. Accordingly, in the case of these items, participants spent the highest mean time and made the most effort, while the mean query-type adequacy score is the highest. In the case of items where subjects were expected to use the *Writing aid* option, there was considerable difference between individual questions in terms of the number of correct answers provided. Finally, the two questionnaire items where we considered as optimal access paths both the *Dictionary module* and the *Writing aid* option, are among the questions with the lowest number of correct answers.

Table 3 provides a summary of search options used to obtain correct answers in the case of each questionnaire item. The highlighted squares represent the optimal query type in each case, which was in fact the most frequently used search option for the majority of questionnaire items. However, it can be seen that the advanced search options were generally under-used, especially the *Writing aid*.

As for effectiveness and efficiency according to user profiles, the group of translation students performed best since they obtained the highest number of correct answers (mean = 10.6, SD = 1.67), and took the least time (mean = 28:55, SD = 11:32) to complete the queries. The native university students performed slightly better

concerning the number of correct answers (mean = 10.38, SD = 2.8) than the foreign university students (mean = 10.0, SD = 4.14), whereas the group of native Spanish language teachers seemed to have the greatest difficulty in using the interface (mean = 6.8, SD = 4.21).

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
Dictionary module	10	8	9	7		12	5		5	12	6	9	12
What does it mean?		17		16	1						14		10
Writing aid	6		12			7				3	1	2	
Direct search							16		13				
Inverse search					17			19				1	
TOTAL	16	25	21	23	18	19	21	19	18	15	22	11	22

Table 3: Summary of the number of correct answers generated using each query type per question

Information on user satisfaction was collected in a post-test questionnaire. In addition, following each query during the usability test, participants were asked to assess its difficulty on a 1–5 Likert-type scale. In the post-test questionnaire, participants were asked whether they had used DiCE before and whether they used it frequently. The answers to these questions reveal that none of the participants had substantial experience with the dictionary. In contrast, 20 participants answered “yes” when asked whether they would use the dictionary in the future, while the remaining six said “maybe”. Finally, when asked whether they would recommend the dictionary to others, 20 informants said “yes”, three said “maybe”, and the remaining three participants said that they would recommend it but it is not easy to use, or they would recommend only the simpler features. In conclusion, participants’ answers reveal a clear positive attitude towards DiCE, although some have reservations about its ease of use. This last point is also apparent if we observe the difficulty score assigned to questionnaire items. The mean difficulty score assigned by participants is 2.65 (SD = 0.77).

5. Discussion

As we have seen, the items of the usability questionnaire were designed in a way that they encourage users to experiment with the different advanced search options available in the DiCE web interface. However, as the results presented above suggest, subjects most frequently used the *Dictionary module*. The reasons for this are twofold. On the one hand, this access path is offered by default in the web interface, and, in addition, it assists in retrieval of the correct answer in the case of most questionnaire items; consequently when participants managed to find the required information using this feature, they did not subsequently employ any advanced

search options for the task. On the other hand, this module provides a type of access similar to paper dictionaries, which may therefore be more familiar to users. Among the advanced search options, the most frequently and successfully used query type was *What does it mean?*. We believe that this can be accounted for by the way dictionaries are most commonly used: users tend to check a given lexical item (either for its meaning or spelling), but they generally do not search for how to express a specific meaning. Also, note that two of the four questions where this option was indicated as an optimal query type explicitly asked about the *meaning* of collocations, which might have served as a clue for users as to which search option to choose.

A qualitative assessment of individual search options has allowed us to explore what details in particular were problematic from the user's point of view. Most identified problem areas can be referred to as *content related problems*, given that they reflect the informants' difficulties in interpreting the dictionary content and the presentation of lexicographic data. The most prominent of these was a lack of familiarity with the notion of collocations and the specific terminology applied in DiCE. Subjects tended to confuse the elements of a collocation (base and collocate) leading to difficulties in using a number of search options. For instance, in the *What does it mean?* search option the search form requires introducing the base and the collocate in individual search boxes; nevertheless, one participant typed a whole collocation string in the box corresponding to the collocate, while others interchanged the two elements of the collocation instead of writing them in the corresponding search boxes. We also noticed that participants tended to confuse the *Direct search* and the *Inverse search* options, which might be a consequence of the fact that both search forms require the introduction of an element of a collocation (the base or the collocate, respectively).

In DiCE, the approximate meaning of collocations is described via a semantic gloss, which some participants tended to confuse with the collocates themselves. For example, when accessing the lexical entry of a base through the *Dictionary module*, collocates are grouped in such a way that the user is provided with a list of gloss tabs, which must be opened to access the collocates. After using the *Dictionary module*, some participants included glosses in their answers, listing them together with collocates, while a few subjects only listed the glosses themselves, which suggests that they were unaware of the need to open the gloss tab to visualize the collocates. We also noticed a few cases when participants typed a semantic gloss in a search box corresponding to the collocate, for instance, in the *What does it mean?* search option.

Some participants proved to be unfamiliar with the more general concepts of word form and lemma. In the case of *Inverse search*, when introducing a collocate in the search box, users can choose between searching for the exact word form (e.g. the feminine or the masculine form of an adjective) or the lemma, the former being the default search option. A number of queries reveal that the distinction between lemma and word form was not familiar to a few participants. In addition, we also noticed that participants experienced some difficulty in identifying and distinguishing lexical

units. In fact the two questionnaire items for which we obtained the lowest number of correct answers involved identifying a particular lexical unit, and providing its collocates, on the basis of example sentences.

It follows from the above considerations that although the user-friendliness of the DiCE interface can clearly be improved, our results also imply the importance of users' reference skills. We have seen that the participants of our experiment lacked some of the knowledge necessary to successfully use the more advanced functions of the dictionary. This claim is supported by the comparison of the performance of the participant groups of varying user profiles. We have seen that the group of translators performed best, which may be a result of the fact that they may be more used to dealing with different lexical tools. The group of language teachers displayed the poorest results, though it should be noted that they incidentally also belong to an older age group than the rest of the participants, and probably have less experience in using web interfaces in general. In any case, we believe that a demonstration of the DiCE website or the use of familiarization activities prior to the experiment, as in the case of Hamel (2012), would have resulted in a considerably better test performance of most participants. In fact, it should be noted that the only informant who claimed to have completed the web tutorial prior to the experiment itself, performed substantially better on the test than the rest of the participants.

6. Conclusion and future work

This paper has described a usability study of the DiCE web interface. The results of the test above all point to the importance of user familiarization with the concepts used by the dictionary. On the one hand, we believe that a number of changes to the current design can considerably improve dictionary usability. These include a more consistent exemplification of the content to be introduced in each search box, a clear indication of obligatory search boxes and filters, as well as the enhancement of the visibility and distinguishability of navigation aids, e.g. semantic glosses and buttons that allow expansion and contraction information to be shown on the screen. On the other hand, we think that, in order to obtain a clearer picture of the usability of DiCE, future research should better control for reference skills of participants, and include familiarization tasks. Finally, we would like to emphasize that the methodology applied in this experiment implies that the test can be completed on participants' home computers, which considerably facilitates data collection and, therefore, may be of interest for future user experiments.

7. Acknowledgements

This work was supported by the Spanish Ministry of Economy and Competitiveness (MINECO) and the FEDER Funds of the European Commission under the contract number FFI2011-30219-C02-01, as well as the Spanish Ministry of Education under the FPU grant AP2010-4334. We would also like to thank Miguel Ramos Naveira for his help with retrieving the log files.

8. References

- Alonso Ramos, M. (2004). *Diccionario de colocaciones del español*. Accessed at: <http://www.dicesp.com>.
- Alonso Ramos, M., Nishikawa, A., & Vincze, O. (2010). DiCE in the web: An online Spanish collocation dictionary. In S. Granger & M. Paquot (eds.) *eLexicography in the 21st century: New challenges, new applications. Proceedings of eLex 2009*. Presses Louvain-la-Neuve: Universitaires de Louvain, pp. 367–368.
- Hamel, M.-J. (2012). Testing aspects of the usability of an online learner dictionary prototype: a product- and process-oriented study. *Computer Assisted Language Learning*, 25(4), pp. 339–365.
- Heid, U. (2011). Electronic dictionaries as tools: Towards an assessment of usability. In P. A. Fuentes-Oliveira & H. Bergenholtz (eds.) *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. New York: Continuum, pp. 287–304.
- Heid, U., & Zimmermann, J. T. (2012). Usability testing as a tool for e-dictionary design: collocations as a case in point. In R. V. Fjeld & J. M. Torjusen (eds.) *Proceedings of Euralex 2012*. Oslo, pp. 661–671.
- Jousse, A.-L., L'Homme, M.-C., Leroyer, P., & Robichaud, B. (2011). Presenting collocates in a dictionary of computing and the Internet according to user needs. In I. Boguslavsky & L. Wanner (eds.) *Proceedings of the 5th International Conference on Meaning-Text Theory*. Barcelona, pp. 134–144.
- Mel'čuk, I., Clas, A., & Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve: Duculot.
- Nielsen, J. (1993). *Usability engineering*. San Francisco: Morgan Kaufman.
- Vincze, O., Mosqueira, E., & Alonso Ramos, M. (2011). An online collocation dictionary of Spanish. In I. Boguslavsky & L. Wanner (eds.) *Proceedings of the 5th International Conference on Meaning-Text Theory*. Barcelona, pp. 275–286.