LEXMATH - A Tool for the Study of Available Lexicon in Mathematics

Pedro Salcedo Lagos[1], María del Valle[1], Ricardo Contreras Arriagada[2](✉), and M. Angélica Pinninghoff[2]

[1] Department of Research and Educational Informatics, University of Concepción, Chile
[2] Department of Computer Science University of Concepción, Chile
{psalcedo,mdelvall,rcontrer,mpinning}@udec.cl

Abstract. To have a good command of the language, not only allows us to write effectively a message, but it also helps us to understand what the other person wants to communicate. Students are in the process of learning to master the language. In particular, as they increase their level of education, they find a series of new terms (lexicon), which need to be understood using appropriate strategies, not only in the subjects of humanistic education, but also in scientific training and especially in mathematics. In the study of the lexicon, from their different points of view, the use lexical statistics gains in importance, which gives us the possibility of quantifying the lexical units of a language using different statistical indicators. Through this work the LexMath tool will be described. This software was developed for the purpose of having an online software on the Internet, for measuring the vocabulary of a particular group in a specific area and present easily the main indicators for the study of the latent lexicon of a given population, and the graphs which enable the study of semantic relations formed between them.

Keywords: Lexicon · Hypermedia · Mathematics · Graphs · Language

1 Introduction

The lexicon is the vocabulary of a particular language, field, social class, person or specific domain of knowledge. The available lexicon for a person is the set of words he knows and uses. Lexical availability refers to the field of research that consists in the recollection and analysis of the available lexicon in a specific community when dealing with a specific topic [12], [13].

Lexical availability studies began in the 50s associated to the French linguistic with the research of R. Michea [20]. The objective of the research was to make easy, for immigrants in France, the learning of the French language. To collect the available lexicon in the community, some points of interest were considered, and around these points, individuals considered in the sample should generate a list of lexical units. Subsequently, in different countries different researchers

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2015, Part II, LNCS 9108, pp. 11–19, 2015.
DOI: 10.1007/978-3-319-18833-1_2

have developed analogous efforts, [21] in Canada, [11] in England, [19], [22] in Puerto Rico, [5] in Spain. In Chile, specifically at the University of Concepción, it is possible to highlight the works of Mena and Echeverría [12], [13].

The central question is, why it is important to study lexical availability in people? The answer is that different studies indicate that through the analysis of lexical availability it is possible to measure the level of knowledge for a specific topic. In other words, the level of knowledge is related to the lexicon a person can handle. The available lexicon let us know the vocabulary a particular society uses for referring to a determined field, allowing researchers to establish the way in which a community understand a concept, to establish weaknesses, and other attributes related to the way in which people generate relationships between words and concepts.

To establish, through the lexicon, the level of knowledge that a person has about a particular topic, it is necessary to measure the lexicon. The way in which it is accomplished is by asking to a group that belong to a community (a class in a school, for instance), to write all the words associated to a topic they can remember, during a short time interval, let us say two minutes. Then, some statistic analysis is carried out for getting the number of different words, the frequency of appearance of words, distribution or position in which words appear, and so on.

Because of the huge volume of information, at the lexicon level and from the point of view of statistics, it is reasonable to develop a computational platform for supporting research in the field, for both, information gathering and processing of the collected information. The core idea of this work is to introduce the platform operation description. The platform, specifically developed for this purpose is called LexMath.

This article is structured as follows; the first section is made up of the present introduction; the second section describes the problem we are dealing with, the third section describe the LexMath platform; the fourth section describes how the platform was used, and the final section shows the conclusions of the work.

2 The Problem

Talking about the process of concepts learning, Baddeley in his book "Memoria" [6] says that the concepts are learned in similar contexts, that most of material that a lecturer teach is not transferred to students, due to a lack of general knowledge, to the misunderstanding on the applicability of the particular object under consideration. In other words, depending on the amount of known terms or words in a close relationship with the concept the lecturer is teaching, the more easy is to acquire it.

Besides that, there are a set of terms related to the concept of lexicon, that are used by researchers:

- Usual or basic lexicon. Subset of a vocabulary used in everyday life, that in characterized for a high degree of use, appearing frequently in every lexical interaction, and that is independent of the topic under consideration.

- Mental lexicon. Is the capability of understanding and using lexical units that allow people to realize an interpretation or to generate units, that were not previously detected, and combine these units with other available units. This mental lexicon represents the individual knowledge as part of the vocabulary a community holds.
- Lexical availability index (LAI). It is a value that indicates the availability degree of a word in an interest center. An interest center is a particular topic that manage a specific vocabulary. The LAI considers the frequency that a word is used in every position in which it is used (named) [12].

To obtain a basic lexicon it is usual to work with texts and to use frequency and dispersion as parameters; the available lexicon is obtained through surveys, by using a stimulus encouraging the individual to update his/her mental lexicon, that in words of Emmorey and Fromkin [14] is the "grammatical component that contains information about the words that are necessary to the speaker". The information of the words is obtained through phonological, morphological, syntactic and semantic information (meaning or conceptual structure [18]). According to Hall [17] words in the mental lexicon are acquired and memorized through speaking, orthography, syntactic frame and concept or word meaning.

To obtain the available lexicon in a community, different interests center are prepared (food, games, professions, etc.). Around these centers of interest centers individuals create lists of lexical units in a limited time; only two minutes for each center of interest. These kind of experiences are found in [24], that described a dictionary containing the available lexicon in Navarra (Spain), or in [1] that exhibit the available lexicon in Jaén (Spain). Similar experiences can be found in Aragón [3], Ceuta [4], Las Palmas de Gran Canaria [8], Río Piedras in Puerto Rico [9], and in the community of Valencia [10].

In mathematics, the closest reference is the work of Echeverría et al. [13] and [15]; the first apply a survey of lexical availability to students and lecturers in Mathematical Engineering career (University of Concepción, Chile). The aim of this survey was to identify which is the available lexicon for these students in interests centers associated with topics that students are expected to know (calculus, algebra, statistics, physics and geometry). In [15] the lexicon available in a set of 1557 high school students is quantified and described, showing that there is an increment in the available lexicon as student increase their formative level.

It is important to note that the mechanism employed to obtain different lexicons by using conventional approaches, writing words in a paper in less that two minutes, requires important resources: surveyors, paper, time, among others. In a research project Fondecyt [16] this method has been applied to a sample of 1557 students on seven centers of interest, taking a set of 60000 words for analysis, after correcting some orthographical errors, correcting some grammatical errors, and generating rules for transforming similar words, like diminutives [23]. By considering all the previous difficulties, it was taken the decision of developing a tool for automating this work.

3 LexMath

LexMath is created to establish an innovation in the way in which the available lexicon for a group of students is collected (see Figure 1 and 2). LexMath proposes methods for increasing the lexicon through an adaptive hypermedia; offering adequate tools for a teacher in such a way that it can be considered as a useful mechanism supporting the teaching strategies, in particular when working with Numbers, Geometry, Algebra, Data and Randomness, which are considered in the subtopic of mathematica during the normal cycle of formal learning.

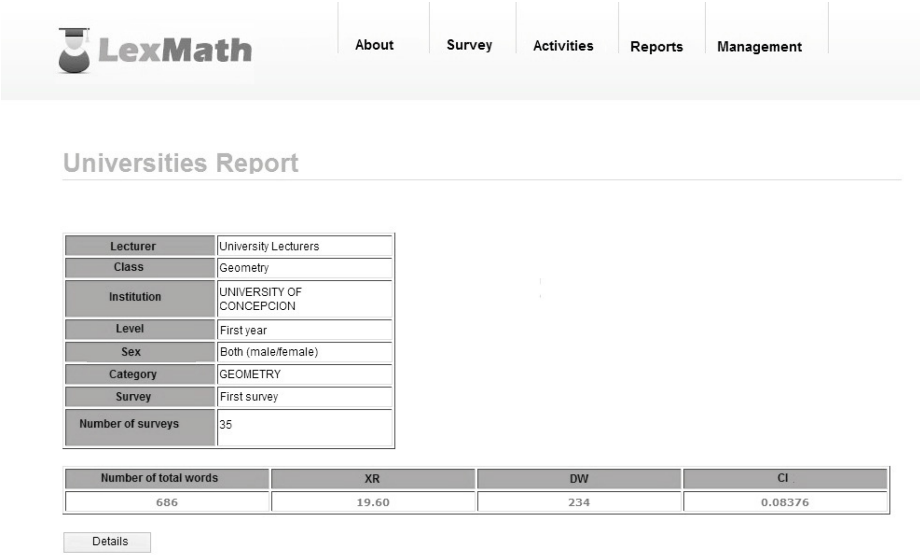

Fig. 1. Available lexical report

Figure 1 shows a screen with results from LexMath. In the top border different buttons make visible the platform functionalities. *About* describes LexMath; *Survey* allows for the students to write the words that they are thinking, related to an interest center, in a time interval of two minutes; *Activities* proposes the activities that are more suitable to the lexicon the student holds (in adaptive way); *Reports* permits to display a table for showing the lexical availability (list of words ordered according to the LAI index), related to a specific question. *Management* (a restricted option) allows to add, modify or delete information in the platform. The center of the screen displays information about a specific center of interest (in this case Geometry) and some characteristics about the group being tested. The bottom part of the screen shows results: total of words, that is the number of collected words; XR is the average of words, i.e., the number of total registered words divided into the number of individuals in the sample, DW is the number of different words collected; and CI is the Cohesion index, that represents the similarity of words appearing in the survey.

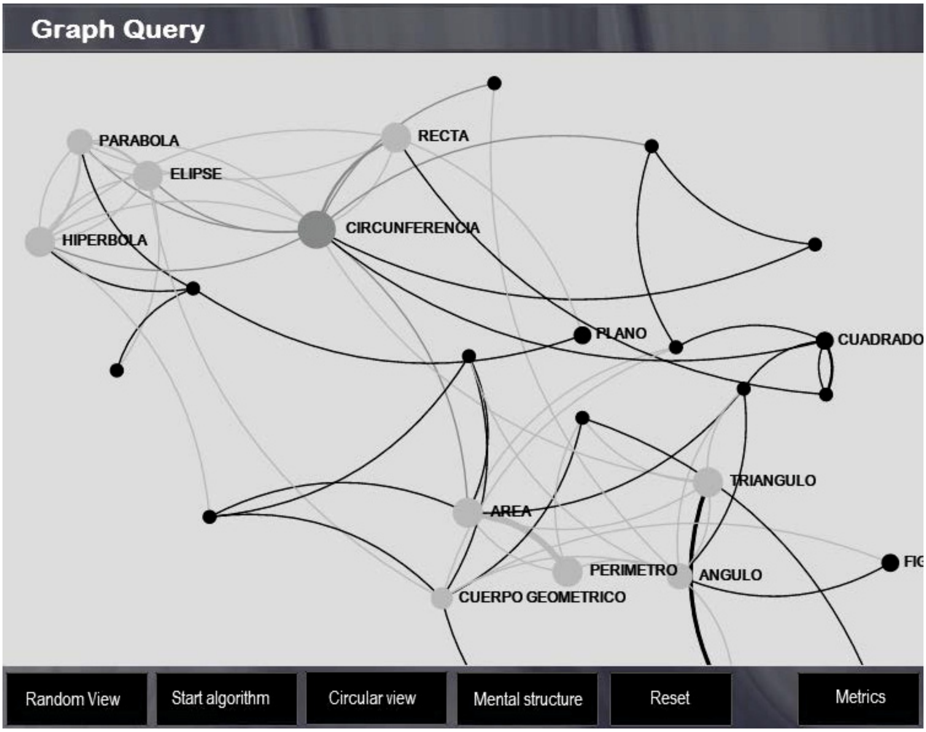

Fig. 2. Semantic structure as a graph

LexMath allows to visualize semantic relationships, taking into account the frequency of different sequences, when testing lexical availability (see Figure 2). Nodes in the graph are the words, the size of each node increases as increases the number of times that a word appears in the answers of a group. Edges indicate the sequence in which words are written. The edge's thickness increases as the words connected by these edges are written in the same order. It is possible that in a highly connected network visualization is not an easy task. Due to this issue, LexMath allows to delete relationships that are not relevant (i.e., relationships that appear with a low frequency). It is useful for researchers if they are interested in some specific relationships; in other words, they can focus only on strong relationships.

Figure 2 is a screen with a graph that shows semantic relationships related to Geometry. Additionally, the platform allows to show the graph in different schemas, as indicated through the buttons appearing on the bottom part of the image. *Random View* generates a view in which nodes are randomly distributed;

Start algorithm allows nodes distribution through an algorithm for obtaining an improved view of the graph; *Circular view* display nodes in a circumference and distributes edges inside the circumference for having a different graph visualization; *Mental structure* allows to hide less frequent nodes (lower weight nodes) displaying only the most significant clusters; *Reset* recover the initial display of the graph; and finally *Metrics* is used to show the values of the main graph indexes, such as degree, degree centrality, graph density, among others.

The Adaptive Hypermedia is developed in terms of the model presented in Figure 3, which contains four well defined components. These components interact to offer the user a hypermedia according to their lexical requirements.

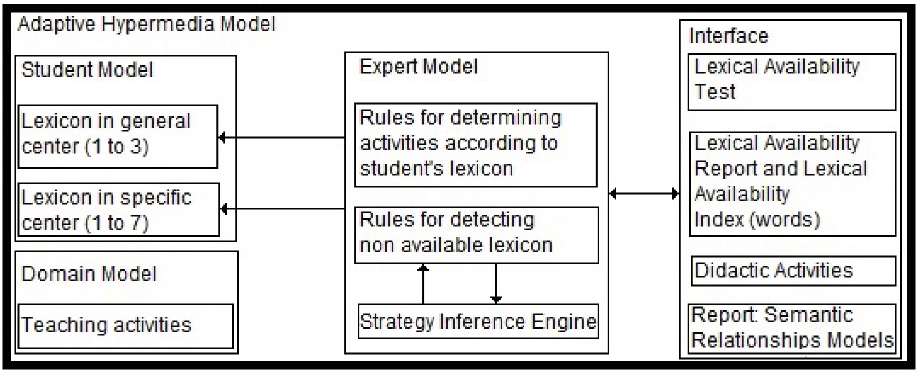

Fig. 3. Semantic relationships in LexMath

- Student model component: contains the databases that hold the lexical availability for each student, in general and specific centers of interest.
- Domain model component: contains the databases associated to different media and activities, according to a didactic proposal, for increasing the available lexicon in different centers of interest, and the ideal lexicon to be obtained by applying the same survey to the teachers.
- Expert model component: contains all the rules that are necessary to establish the general lexicon for a student and the non available lexicon. Additionally it contains the inference engine, which is in charge of determining the adequate activities in terms of the student model and the domain model.
- Interface: This is the component that allows to collect the lexicon, through personalized activities and to generate different reports, including statistical reports.

For generating the graph shown in Figure 2, we use the Script Gexfjs.js (http://brandonaaron.net), open source for academic use, supported by the Gephi technology [7].

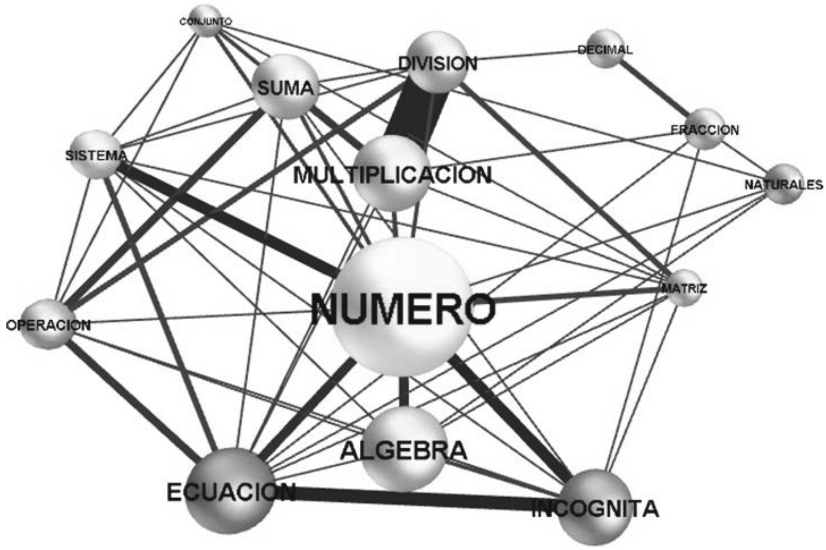

Fig. 4. Semantic relationships for the center of interest Algebra

4 LexMath: Application

During the development of LexMath [23], the use of resources was an important issue, and it was established a strategy for applying the tests in exactly 2 minutes through Internet, finding orthographical errors and grammatical errors, storing all the information in a database, including psychosocial characteristics for every student considered in the survey. Additionally, the system allows to access some tools that automatically generate statistics views, and display semantic networks for visualizing the most relevant semantic structures (lexical profiles or mental models).

In Concepción (Chile), it was conducted a research about lexical analysis supported by LexMath. The hypothesis is that exist differences in the vocabulary associated to the field of mathematics, according to socio-economical level different institutions present. The goal of this research was to analyze the lexicon and semantic networks on a sample of high school students.

Surveys were applied to 14 educational institutions, representing a wide spectrum in socio-economical level. This sample considered a total population of 1557 students, and the aim of the surveys was to collect the lexicon students presented around seven centers of interest.

In Figure 4 is shown an example of a graph generated by using LexMath, it is possible to identify the semantic relationships established when cosidering a particular center of interest: Algebra. In the set of 14 nodes that appears on the graph, it seems that the most important words, in terms of number of relationships and number of times that they were mentioned by students, are NUMERO (number), ECUACION (equation) and INCOGNITA (unknown). The first two words hold eleven of the thirteen possible links with the rest of the nodes, while the third (INCOGNITA) holds eight of these links.

5 Conclusions

The main observed result is that LexMath allows to carry, on line, most of necessary tasks for realizing the lexical analysis for a specific group, faster than the traditional procedure that uses paper, displaying in a few seconds statistics and graphics. The mentioned research hypothesis was confirmed, in the sense that exist differences in semantic networks for identical center of interest, when considering different social-economical characteristics in students. It is important to note that students with a lower social-economical level show less nodes and clusters, i.e., present visible weaknesses when dealing with specific concepts. No doubt, it is a very important issue that should be addressed in future initiatives concerned with quality and equality opportunities.

Acknowledgement. This work is partially supported by the Chilean National Fund for Scientific and Technological Development, FONDECYT, through research project number 1140457.

References

1. Lara, A.:: El léxico disponible de los estudiantes de Jaén, Jaén University (2006)
2. Aguilar, M., Navarro, J., López, J., Alcalde, C.: Pensamiento formal y resolución de problemas matemáticos. Psicothema 14(2), 382–386 (2002), Revised in psicothema.com/pdf/736.pdf
3. Arnal, M. L.(coord.): Léxico disponible de Aragón, Zaragoza: Libros Pórtico (2004)
4. Ayora, C.: Disponibilidad léxica en Ceuta: aspectos sociolinguísticos, Cádiz University (2004)
5. Azurmendi, M.J.: Elaboración de un modelo para la descripción sociolinguística del bilinguismo y su aplicación parcial a la comarca de San Sebastián. San Sebastián: Caja de Ahorros Provincial de Guipúzcoa (1983)
6. Baddeley, A., Eysenck, M., Anderson, M.: Memoria. Alianza Editorial, Madrid (2010)
7. Bastian, M., Heymann, S., Jacomy, M.: Gephi: an open source software for exploring and manipulating networks. In: International AAAI Conference on Weblogs and Social Media (2009)
8. Bellón: Léxico disponible de la provincia de Córdoba, Las Palmas de Gran Canaria, Doctoral Thesis (2003)
9. Butrón, G.: El léxico disponible: índices de disponibilidad, Río Piedras: Universidad de Puerto Rico (Doctoral Thesis, unpublished) (1987)
10. Casanova, M.: La disponibilidad léxica en la Comunidad Valenciana. In: Blas, J.L., Casanovay, M., Velando, M.(coords.): Discurso y sociedad. Contribuciones al Estudio de la lengua en Contexto Social, Castellón de la Plana: Universidad Jaume I, 737–752 (2006)
11. Dimitrijevic, N.: Lexical Availability. Groos, Heidelberg (1969)

12. Echeverría, M.: Crecimiento de la disponibilidad léxica en estudiantes chilenos de nivel básico y medio. En: Morales, H.L.(ed.) La enseñanza del español como lengua materna, pp. 61–78. Universidad de Puerto Rico, Río Piedras (1991)
13. Echeverría, M., Urzua, P., Sáez, K.: Disponibilidad Léxica Matemática. Análisis cuantitativo y cualitativo. In: Revista de Linguística Teórica y Aplicada Concepción (Chile), II Sem. 2006, vol. 44 (2), pp. 59-76. Universidad de Concepción, Chile (2006)
14. Emmorey, K., Fromkin, V.: The Mental Lexicon. In: Newmeyer, F. (ed.) Linguistics: The Cambridge Survey, vol. III. CUP, Cambridge (1988)
15. Ferreira, A., Salcedo, P., del Valle: Estudio de disponibilidad léxica en el ámbito de las matemáticas. Estudios Filológicos, Universidad Austral de Chile, Valdivia 54, 69–84 (2014)
16. Pedro Salcedo L., Maria del Valle L., Anita Ferreira C., Nail, O.: Project: Fondecyt 1120911. Disponibilidad Léxica Matemática en Estudiantes de Enseñanza Media y su Aplicación en Hipermedios Adaptativos, Investigadores (2012, 2013), http://www.lexmath.com
17. Hall, J.H.: Morphology and Mind. Routledge, Londres (1992)
18. Jackendoff, R.: Semantic structures. MIT Press, Cambridge (1990)
19. López Morales, H.: Léxico disponible de Puerto Rico. Arco/Libros, Madrid (1999)
20. Michea, R.: Mots fréquents et mots disponibles. Un aspect nouveau de la statistique du langage. en Les Langues Modernes 47, 338–344 (1953)
21. Mackey, W.C.: Le vocabulaire disponible du franaise, 2 vols. Paris-Bruxelles-Montreal (1971)
22. Román, B.: Disponibilidad léxica de los escolares de primero, tercero y quinto grados del distrito escolar de Dorado (Master Thesis). Universidad de Puerto Rico, Recinto de Río Piedras (1985)
23. Salcedo, P., Nail, O., Arzola, C.: "Análisis de Relaciones Semánticas del Léxico Disponible en Matemáticas en un Hipermedio Adaptativo". Nuevas Ideas en Informática Educativa. In: Actas del Congreso Internacional de Informática Educativa. Santiago de Chile, vol.8, pp. 154–158 (Diciembre2012)
24. Saralegui, C., Tabernero, C. (2008). Aportaciones al proyecto Panhispánico de Léxico Disponible: Navarra. In: XXXVII Simposio Internacional de la Sociedad Española de Linguística (SEL), Navarra (2008) ISBN: 84-8081-053-X
25. Urzúa, P., Sáez, K., Echeverría, M.: Disponibilidad léxica matemática, análisis cuantitativo y cualitativo. Revista de Linguística Teórica y Aplicada (RLA) 44(2) (2006)

Springer

http://www.springer.com/978-3-319-18832-4

Bioinspired Computation in Artificial Systems
International Work-Conference on the Interplay
Between Natural and Artificial Computation, IWINAC
2015, Elche, Spain, June 1-5, 2015, Proceedings, Part II
Ferrández Vicente, J.M.; Álvarez-Sánchez, J.R.; de la Paz
López, F.; Toledo-Moreo, F.J.; Adeli, H. (Eds.)
2015, XXIV, 462 p. 177 illus., Softcover
ISBN: 978-3-319-18832-4