

## A semantic and pragmatic classification of the vocabulary of educational psychology research articles

Andrea Silvana Garofolo  
Universidad Nacional de Río Cuarto, Facultad de Ciencias Humanas  
[agarofolo@hum.unrc.edu.ar](mailto:agarofolo@hum.unrc.edu.ar)

(Received: June 20, 2022; accepted: September 21, 2022)

### Abstract

This paper provides a classification of the vocabulary of the educational psychology research article based on semantic and pragmatic criteria. This study provides a description of a corpus of research articles in the field of educational psychology, focusing particularly on the coverage and frequency of the lexical items, based on the General Service List (GSL) and the Academic Word List (AWL). We analyzed and reclassified the most frequent words of the corpus according to semantic and pragmatic criteria, which revealed that a large number of words coming from the GSL and the AWL were not used with general or academic meaning in the field of educational psychology. This semantic and pragmatic classification allowed the compilation of a list of academic and technical words representative of the genre and discipline.

*Keywords:* Technical and academic vocabulary; semantic and pragmatic criteria; research article; educational psychology

### Resumen

Este artículo propone una clasificación del vocabulario del artículo de investigación de psicología educativa según criterios semánticos y pragmáticos. Este estudio ofrece una descripción de un corpus de artículos de investigación del área de la psicología educativa, centrándose particularmente en la cobertura y frecuencia de los recursos léxicos, en base a las listas General Service List (GSL) y Academic Word List (AWL). Analizamos y reclasificamos las palabras más frecuentes del corpus según criterios semánticos y pragmáticos, lo cual reveló que una alta proporción de palabras provenientes de la GSL y la AWL no se usan con significado general o académico en el área de la psicología educativa. Esta clasificación semántica y pragmática permitió la construcción de una lista de palabras académicas y técnicas representativas del género y la disciplina.

*Palabras claves:* Vocabulario técnico y académico; criterios semánticos y pragmáticos; artículo de investigación; psicología educativa

## Introduction

Vocabulary research in the field of English for Academic Purposes (EAP) has been prompted by the need to facilitate non-Anglophone learners' access to the vocabulary considered necessary to function in academic studies. Attempting to satisfy this need, many authors have centered their attention on the construction of word lists designed to help teachers in the selection of useful words for learners. The best known and most widely used lists have been the General Service List (GSL) (West, 1953) and the Academic Word List (AWL) (Coxhead, 2000). Recently, however, these lists were strongly scrutinized, and two new lists, the New General Service List and the New Academic Word List, were produced to redress deficiencies, such as outdatedness and inconsistent selection criteria in the GSL (Brezina & Gablasova, 2013), and corpus size and the use of word families as units of word count in the AWL (Gardner & Davis, 2013).

These four lists appear to have been particularly designed to be functional in general contexts, such as that of English for General Academic Purposes (EGAP) (Hyland, 2006). For example, the AWL was designed to provide general academic vocabulary useful for non-Anglophone students learning English in English-speaking countries. Practitioners in such contexts need to teach a wide vocabulary because the learners generally have distinct fields of interest and need to function academically using a wide variety of genres. Such is the case of the students described by Coxhead (2011) in her New Zealand situation, and that of a large number of non-Anglophone students studying in English-speaking countries (Garofolo, 2011; Hyon, 1996; Swales, 1990).

There are other situations where the use of lists of general vocabulary does not seem to be effective for students, as it occurs in contexts where English is an auxiliary language and students generally have little proficiency in English. These students, in contrast with those in English-speaking countries, have similar fields of interest and need to use a limited - sometimes only one - set of genres. Such is the case of Ward's (2009) engineering students in Thailand, who need to study engineering from textbooks in English. For these contexts, researchers have suggested teaching vocabulary, focusing on their highly specific fields of interest and providing only the words necessary to function in their particular area of knowledge (Garofolo, 2011; Garofolo et al., 2009; Durrant, 2016; Hyland & Tse, 2007; Ward, 2009).

The need for specificity in vocabulary teaching was foregrounded in a timely and highly influential article by Hyland and Tse (2007), in which the authors were critical of the use of general lists for vocabulary teaching, specifically in relation to academic vocabulary and Coxhead's (2000) AWL. Using a multi-disciplinary corpus, they demonstrated that Coxhead's AWL provided more words than those that students of specific fields needed, while many words that they did need were not included in the list. On these grounds, Hyland and Tse suggested building word lists to meet the students' highly specific needs in their particular fields of study. This view exerted a strong influence on researchers, who turned to focus on field specific vocabulary, producing many discipline-specific lists, for example an agriculture list (Garofolo et al., 2009), an applied linguistics list (Vongpumivitch, Huang & Chang, 2009), a financial list (Li & Qian, 2010), and several medical lists (Chen & Ge, 2007; Lei & Liu, 2016; Wang, Liang, & Ge, 2008).

Despite the large number of publications on word lists, many problems still persist. For example, the frequency and range criteria used in the construction of the AWL do not seem to guarantee the semantic status of the words selected. In this respect, Hancioglu, Neufeld and Eldridge (2008) observe that there are words in the AWL that should be part of the GSL based on the evidence that they have high frequency in many corpora. Similarly,

Gardner (2013) points out that there are words from the GSL that were not included in the AWL “even though such words have major academic meanings” (Gardner, 2013, p.5).

It can also be observed that in studies using small specific corpora, the academic lists built only on the basis of frequency and range tend to include many words that do not fit in the category “academic”, thus revealing some inconsistencies. Specifically, in the lists of recent small corpus vocabulary studies, some of the ten most frequent academic words identified are, in fact, technical words. Some examples in the field of medicine are *cell, muscular, clinic, protein, tissue, gene, therapy, cancer* in Chen and Ge’s list (2007); *cell, muscular, clinic, protein, therapy, cancer* in Wang, Liang and Ge’s list (2008); and *abdominal, albumin, artery, clinic, clinician, chest, patient, pulmonary* in Lei and Liu’s list (2016). The same phenomenon could be observed in lists produced in other fields, such as the words *income, panel, entity, currency* in Li and Qian’s (2010) financial academic list, and the words *client* and *western* in Gardner and Davies’s (2013) New Academic Word List.

There are also different definitions of technical vocabulary. One is the widely known definition by Nation (2001), which considers that technical words are those items which are specific to the disciplines. This vocabulary was identified quantitatively as being made up of the words that remained after the exclusion of the words in the GSL and the AWL, and the low frequency words. With this selection criterion only 5% of the words of a particular text was considered to be technical. This criterion, however, was revised by Chung and Nation (2004) using semantic criteria, and the authors concluded that the coverage of technical words could be around 30% of a text, largely exceeding the percentage calculated by Nation. As Muñoz (2015) states, the use of frequency criteria may be “misleading” since quantitative procedures for analyzing vocabulary may not allow to see the real meaning of a word in different contexts. Besides, frequency based-descriptions generally show that a large proportion of the subject-related vocabulary in a corpus are general words that have a specialized meaning in specific domains, or words which are used in different disciplines with different meanings (Coxhead and Nation, 2001). These conflicts in vocabulary analysis suggest the need to use other criteria, in addition to word frequency analysis, to identify and study the actual meanings of words as used in specific contexts. The use of semantic and pragmatic criteria represents a more suitable approach to study and classify the vocabulary of specific genres and disciplines and develop word lists to meet students’ needs in particular fields of study within academic contexts.

### Using semantic and pragmatic classification criteria

An approach to identify the meaning of words involves the use of semantic and pragmatic criteria as proposed by Cabré (1999), in the field of terminology, and Pearson (1998). These criteria are particularly relevant to study and classify lexical words, specifically those with technical and academic meaning. As Cabré states, technical words are those items which refer to entities, properties or processes that are related to the subject matter of a field. A technical word, or a term, is a unit described by a set of systematic linguistic characteristics with the property of referring to an element in reality used in a special domain (Cabré, 1999). Semantically, a term can be classified according to the type of concept that it designates, such as objects, processes and actions, properties, states and qualities, and relationships. As Cabré proposes, based on the theory of terminology, a word becomes a term when it is related to all the other terms that form part of the same subject field or discipline, constituting a conceptual field. The concepts that these terms designate are connected to each other through logical relationships (e.g., hyponymy and coordination) and ontological relationships (e.g., chain relations). The combination of these relations determines the hierarchical structure of a subject field (Cabré, 1999).

From a pragmatic perspective, as Pearson (1998) claims, the notion of communicative setting is a factor which helps determine whether words are being used with technical meaning in a text. Pragmatically, terms are used by professionals to refer to specialized topics in a particular subject field in specific communicative situations, as in the case of research articles. Of the different communicative settings that Pearson identifies, the setting expert-expert communication is the most relevant for the analysis of vocabulary in this study. In this setting, writer and reader or speaker and hearer, are expected to have the same or a similar level of expertise. They share a common language and, when they use certain words, each of the participants understands what they mean. As stated by Pearson, “this language differs from general language in that specific meanings have been assigned to the language used, and these have been defined prior to the communication act by an external authority” (p. 36), such as a specialized dictionary. This expert-expert communication takes place in publications in learned journals, academic books, research reports and legal documents, among others.

Meyer (1997) also proposes a pragmatic approach to unveil the role of lexical items that are not subject-oriented. As this author claims, academic or “non- technical vocabulary refers to entities, properties or processes that are not specific to one field” (p.5) but which can be related to a wide range of disciplinary contexts. Some examples of these words are *data*, *fundamental*, *proposal*, *argue* and *implications*. As defined by Meyer, these words represent the language used by researchers to describe and evaluate the “activities, achievements, accomplishments and mental states” (Meyer, 1997, p. 7) involved in research and communication processes. This approach is particularly relevant for the analysis of lexical words as it may help determine the specific senses and functions that they obtain in the academic context. In a recent study, Muñoz (2015) used both semantic and pragmatic criteria (Cabré, 1999; Meyer, 1997; Pearson, 1998) to classify words of popular science texts of the field of agricultural sciences. These criteria represented a highly useful approach to determine the real meanings and uses of technical and non-technical words and to study the differences between these types of vocabulary (Muñoz, 2015).

### **The present study**

Studying the vocabulary associated with a specific discipline and genre is essential for undergraduate students in reading EAP courses at Universidad Nacional de Río Cuarto (UNRC), Argentina. Particularly, psychology students need to read research articles in English related to educational psychology to write their undergraduate theses. These students generally have some basic knowledge of the language but know little about the specialized vocabulary of the texts they read. In particular, they manifest difficulties in understanding the different meanings and uses of many high-frequency words of research articles of this discipline.

In the present article we analyze the vocabulary of a small corpus of educational psychology research articles, which psychology students at UNRC need to read, and use semantic and pragmatic criteria for the classification of the lexical words with technical content, or terms, and those with academic content. Our results revealed that the most frequent words were used with either academic or technical meaning. The main product of our study is a word list made up of relatively few field-specific words which provide a good coverage.

Our findings will shed light on the specific meanings and uses of the most frequent words of the educational psychology research article (EPRA) and will ultimately contribute to informing pedagogic practices of EAP courses in our university.

### **Materials and methods**

This study followed a quantitative and qualitative corpus-based approach (Scott & Tribble, 2006). We first carried out a quantitative analysis to determine the frequency, distribution and coverage of all the lexical items in the corpus, as performed in previous studies in other disciplines (Garofolo et al., 2009; Chen & Ge, 2007; Lian & Ge, 2008; Vongpumivitch et al., 2009; Wang et al., 2008). We then selected the most frequent lexical words in the corpus and matched them with those coming from the GSL and the AWL in terms of their meaning and use. The words were then reclassified using qualitative criteria, both semantic and pragmatic. The reclassification resulted in a novel list made up of only academic and technical vocabulary of educational psychology.

### The Corpus

A small, specialized corpus of educational psychology (hereafter PsychoCorpus) was built following the criteria proposed by Sinclair (1991; 2005). The corpus was representative of a genre (research article), a field (psychology), and within this, a specific domain (educational psychology). The texts were drawn from the journals *Learning and Instruction* and *Learning and Individual Differences*. These journals were selected considering the needs of psychology students at UNRC, who are expected to read and consult research articles from these journals to write their undergraduate theses. The texts were experimental research articles produced in English-speaking universities, published between 2009 and 2011. The journals had an impact factor of 2.768 and 1.526, respectively (Thomson Reuters Journal Citation Reports 2011 ([www.journals.Elsevier.com](http://www.journals.Elsevier.com))). The corpus comprised 44 articles -22 from each journal- with a size of 235,155 tokens and 10,069 types. Tables, diagrams, numbers, references, acknowledgements, captions and appendices were excluded from the word count. This highly specific corpus was considered representative of the language of research articles in the field of educational psychology.

### Data processing and analysis

#### Quantitative analysis: Description of the corpus

The first stage of the study was a quantitative description of the PsychoCorpus to determine the frequency, distribution and coverage of all the words. The frequency analysis of the corpus (235,155 tokens and 10,068 types) showed the expected distribution, with a small number of highly frequent words in the first three frequency bands and a large number of low frequency words or hapax legomena (Sinclair, 1991) in the last frequency levels. The first 20 types accounted for almost 30% of the words in the corpus and appeared in most of the texts, while the last 6,383 types accounted for only around 5% of the tokens of the corpus and occurred in up to four texts (Table 1).

Table 1. Frequency, coverage and range of types and tokens in the PsychoCorpus (*N*: 235,155)

Token frequency	Types	Cumulative Types	Coverage	Cumulative Coverage	Range (N° of texts)
(> 1000)	20	20	68,302	68,302	44-42
(700-999)	12	32	10,120	78,422	44-35
(300-699)	74	106	31,987	110,409	44-5
100-299	274	380	46,202	156,611	44-1
50-99	329	709	22,650	179,261	35-1
20-49	794	1,503	24,073	203,334	26-1
10-19	908	2,411	12,439	215,773	15-1
5-9	1,274	3,685	8,402	224,175	8-1
1-4	<b>6,383</b>	<b>10,068</b>	<b>10,980</b>	<b>235,155</b>	4-1

Then we determined the coverage of the GSL and AWL by using the Matchlist function of the program WordSmith tools 4.0 (Scott, 2004). As the GSL includes both lexical and grammar words, we separated the GSL words into two sublists: grammar (GSL-GrW) and lexical (GSL-LexW) words. Thus, the words of the corpus were grouped into four sets, GSL-GrW, GSL-LexW, AWL words, and “Other words”, that is the items outside those in the lists.

As our study was concerned with vocabulary, we decided to analyze only the lexical words, excluding the grammar words. The rationale for this decision was the fact that grammar words represent a small stable group of words characterized by their members being grouped in closed categories, their lack of external reference, and their lack of inflections, among others (Halliday, 1985; Stubbs, 1986). These words are highly repeated and cover an important proportion of a text (Biber, Johansson, Leech, Conrad, & Finegan, 2000). In our corpus, the grammar words constituted 40% of the total corpus, but we considered that these words could be excluded since they need special treatment and a teaching methodology independent of lexical words. With this elimination, the size of the corpus was reduced to 140,937 words and allowed us to focus only on the remaining 60% lexical vocabulary.

We then selected a set of frequent lexical words of the corpus, on the basis of two criteria: frequency and range. For this selection, we calculated the median of the corpus wordlist, which was obtained by dividing the total number of tokens (140,937) by two (70,468) to identify the middle point, and the position of the type number where it fell: 319. The median was considered as the most appropriate measure to analyze ranked data with extreme values (Levine & Stephan, 2010). This measure provided a restricted set of high frequency words, 319 words, which were further restricted by range. We considered as high-frequency words only those that occurred above the median, with a frequency of at least 85 times, in a range of at least 10 texts. Finally, the words of the resulting list were again classified by matching them with the words of the GSL and the AWL, thus obtaining lists of the most frequent GSL-LexW and AWL words as well as the list of “Other words”.

#### **Qualitative analysis: High-frequency words**

In the next stage, we reclassified the high-frequency lexical words of the corpus as we could observe that the words identified as GSL-Lex words were not general in the PsychoCorpus but clearly had academic or technical meanings. Similarly, many words identified as AWL words did not have academic meaning in the corpus. Thus, we reclassified them through a qualitative analysis using the semantic and pragmatic categories proposed by Cabré (1999) and Meyer (1997). This analysis was supported by the concord tool of the WordSmith Tools (2004) software, which contributed to determining the meaning of the items in their real context of use.

#### **Categories for lexical analysis: Words with technical and academic meaning**

##### **Words with technical meaning**

The semantic and pragmatic perspectives provided by Cabré (1999) and Pearson (1998) were particularly suitable to identify and reclassify the GSL-Lex, AWL and Other words of the PsychoCorpus with technical meaning. A word was considered to be technical when it represented a concept specific to the discipline of educational psychology and when it was related to other terms that formed part of this domain, constituting a conceptual field. The concepts represented by these words were connected to each other through logical relationships, based on the characteristics that concepts shared in the conceptual field, specifically subordination and coordination, and ontological relationships based on contact or proximity of “beings” to each other, such as chain relations. According to the type of

concepts that these terms designated, the words were analyzed and reclassified as technical following the categories proposed by Cabré (1999):

- (1) Objects and entities which referred mainly to abstract entities in educational contexts such as school subjects and instances of evaluation, for example, *language* and *exam*.
- (2) Processes and actions which took place in educational contexts such as *regulation* and *assessment*.
- (3) Properties, states, and qualities which described objects and processes related to the field of education, for example, *behavioral* and *spatial*.

We also used two other categories proposed by Muñoz (2015) to complement Cabré's (1999) classification. These categories were:

- (4) People, that is, the individuals who participated in processes or operations in the field of education such as *teachers* and *children*.
- (5) Places or settings where educational actions or processes took place, for instance, *class*.

The category Factors (6) emerged from the data obtained in the analysis. It referred to facts, situations and feelings which were involved in or affected the processes and actions occurring in educational contexts, for example, *ability* and *motivation*.

As Muñoz (2015) states, these categories demonstrate that technical meaning is determined semantically by the concepts that words represent and pragmatically by the context in which they are used.

#### **Words with academic meaning**

The pragmatic categories proposed by Meyer (1990, in Coxhead & Nation, 2001) and Meyer (1997) were useful for the identification and classification of words with academic meaning in the PsychoCorpus. Based on these categories, academic words were defined as those items which indicated the writer's linguistic acts in a text, words which referred to entities, properties and processes related to scientific activities, and words which established relationships among the concepts related to the field of educational psychology. The words of the corpus with academic meaning were reclassified according to the categories described below:

#### **(1) The domain of the text:**

- (a) Words which referred to the linguistic acts performed in the texts analyzed. These words indicate "what the authors are doing in their texts and what they ascribe to other authors" (Meyer, 1990 in Coxhead & Nation, 2001, p.5), for example, *indicated* and *demonstrated*.
- (b) Textual deixis included the expressions within an utterance which referred to parts of the discourse that contained the utterance, for example, *respectively*.

#### **(2) Elements of scholarly practice:**

- (a) Vocabulary referring to scientific activities. This group included concepts common in the methodology of science which were related to the stages of the research process such as experimentation, empirical investigation and hypothesis testing among others, for example, *research* and *analysis*. This category also included the vocabulary of statistics which referred to the methods and statistical measures applied in the research process, for instance, *p* (p value) and *SD* (standard deviation).
- (b) Properties of scientific activities: these words referred to the features or characteristics of the activities carried out in the research process, for example, *specific*.
- (c) Evaluations of theories, procedures, methods or equipment and results: this group included words or expressions which judged the value of theories, procedures, methods or equipment, for example, *positive* and *consistent*.

- (d) Classifiers of data: these words referred to facts or information, mainly when they were examined and used to confirm hypotheses or to make decisions in a research process, for instance, *data* and *results*.
- (3) Vocabulary referring to the subject matter of scientific activities**
- (a) Lexical expressions referring to tense, aspect, and modality in the texts: these expressions assigned the scientific activities a point on a time axis.
- Temporal deixis included expressions concerned with the time involved or referred to in an utterance, for instance, *present* and *prior*.
  - Modality: these items allowed the researcher to topicalize the modal character of propositions and permitted a more condensed representation of modalized propositions such as *(be) required*.
  - Event quantifiers: these words referred to the quantity of scientific events or processes, usually expressed by determiners or pronouns, for example, *number*.
- (b) Abstract quantities or quantitative properties of states of affairs or entities: *high* and *small*.
- (c) Classifiers of states of affairs: these words were generally abstract nouns which referred to elements already presented in the discourse. The meaning of these nouns was specified by the anaphoric relation they had with their referents, which could be single words or complex states of affairs previously presented as long elaborations of processes, events or entities. They could act as anaphoric items or as general terms to be elaborated on later in a text, for instance, *approach* and *factors*. These general nouns were not technical words but acquired terminological status when they were combined with words with specialized meaning, as in *behavioural approach* and *factors of academic motivation*.
- (d) Relations between states of affairs. This category included vocabulary related to quantitative changes such as *increase*, causal relations like *associated* and inclusion such as *included*, among others.
- (4) Discourse-organizing vocabulary.** This group included those words which gave cohesion and coherence to the texts. These items structured discourse and signaled meaning relationships among clauses and sentences in the articles analyzed. Some examples are *rather* and *(in) contrast*.

#### **Validation of categories for analysis**

The categories for vocabulary identification and analysis were validated through inter-rater agreement to ensure that they were clearly defined. Inter-rater agreement was obtained by asking an independent rater to classify 30% of the most frequent words into technical and academic words according to the categories described above. The rater was an experienced language teacher from the English Teacher Training College at UNRC, who was familiar with the categories used for vocabulary classification. She was provided with 65 randomly chosen words of the PsychoCorpus: 40 items from the GSL-LexW, 18 items from the AWL and 7 from the list of Other words, and the corpus to study these words in their context of use. The rater analyzed and reclassified the words following the categories of academic and technical vocabulary. The results of the rater's analysis revealed an inter-rater agreement of 97%, which was similar to that in other studies such as Chung and Nation's (2003).

## **Results**

### **Coverage of wordlists in the PsychoCorpus**

To determine the coverage and distribution of the words in the PsychoCorpus we performed a quantitative analysis using the existing lists, namely the GSL, separated in this



study into GSL-GrW and GSL-LexW, and the AWL. The words outside the above lists were grouped as *Other words* in order to carry out further analysis. The words from the GSL provided the highest overall coverage, 70%, with 40% grammar words and over 30% lexical words. The AWL represented more than 14% of the words, and the group of *Other words* covered 15.1% of the corpus (Table 2).

Table 2. Coverage of wordlists in the PsychoCorpus (*N*: 235,155)

Word Lists	Types	Tokens	Coverage
GSL-GrW	225	94,218	40.00%
GSL- LexW	2,796	71,896	30.60%
AWL	1,594	33,670	14.30%
Other words	5,453	35,371	15.10%
<b>Total</b>	<b>10,068</b>	<b>235,155</b>	<b>100.00%</b>

### High frequency words

To analyze the lexical vocabulary, we worked with a manageable set of frequent words. These words were selected from the corpus without grammar words (9,844 types and 140,937 tokens) considering frequency and range, as described in Methods. The identified frequent words conformed a universe of 237 types and 66,794 tokens (Table 3). The ten most frequent items were *students, test, study, self, et, al, performance, achievement, learning* and *group* (See Appendix for the complete list). By matching these words with the GSL-LexW and the AWL, we could observe that the highest number of types came from the GSL-LexW (*students, test, study, self, performance, learning, group, time, also* and *differences*), followed by those of the AWL (*achievement, academic, significant, items, research, motivation, task, participation, strategies* and *variables*) and the list of *Other words* (*cognitive, scores, mathematics, emotions* and *spatial*).

Table 3. Distribution of high frequency words of the corpus in relation to the GSL-Lex, the AWL and Other words (*N*: 66,794)

Word Lists	Types	Tokens	Percentage
GSL-Lex	<b>139</b>	40,351	60.42
AWL	<b>62</b>	16,444	24.62
Other words	<b>36</b>	9,999	14.96
<b>Total</b>	<b>237</b>	<b>66,794</b>	<b>100</b>

### Qualitative classification of the most frequent words of the PsychoCorpus based on semantic and pragmatic criteria

The qualitative analysis of the most frequent words revealed that a large number of words coming from the GSL-LexW and the AWL did not have the meanings suggested by the labels of these lists. A concordance analysis of these words showed that, a considerable number of words from the GSL-LexW, for example *students, test, performance, related* and *suggests*, did not have general meaning, and many words from the AWL, such as *affect, styles, assessment* and *outcomes*, did not have academic meaning in our corpus. Instead, these words were clearly technical in the field as they referred to concepts associated with the field of educational psychology. These findings suggested carefully analyzing each of the most frequent words to determine their actual meaning and use in the corpus. As

described in Methods, we reclassified the words using the categories of academic (Meyer, 1990 in Coxhead & Nation, 2001; Meyer, 1997), and technical vocabulary (Cabr e, 1999).

The qualitative analysis confirmed that none of the most frequent words coming from the GSL-LexW were used with general meaning in the PsychoCorpus; instead, they were used as either academic or technical words. As expected, a good percentage of the words coming from the AWL did have academic meaning. However, it was found that the rest of the words from this list had technical meaning in our corpus. Furthermore, the list of *Other words* only included items with either academic or technical meaning (Table 4).

**Table 4.** Qualitative classification of frequent words in the PsychoCorpus (*N*: 237)

Word Lists	Meanings			Total
	General meaning	Academic meaning	Technical meaning	
GSL-Lex	—	79	60	139
AWL	—	39	23	62
Other words	—	21	15	36
<b>Total</b>	—	139	98	237

#### Words with technical meaning

The semantic analysis showed that 60 words from the GSL-LexW, 23 from the AWL and 15 from the list of *Other words* had, in fact, technical meaning in the PsychoCorpus. Some examples are *students, teachers, school, content, class, interest* (GSL-LexW); *achievement, critical, motivation* (AWL); and *comprehension, cognitive, emotions, mathematics* (Other words). These words were reclassified as technical following the categories proposed by Cabr e (1999), as described in Section 2.2.3. Some examples are provided below:

- (1) Objects and entities in the field of education:
  - (1) *...students' intention to engage further with science content...*
  - (2) *...devote less time and effort to studying mathematics...*
- (2) Processes and actions which took place in educational contexts:
  - (3) *High anxiety was detrimental to students' performance...*
  - (4) *...affecting students' achievements and learning...*
- (3) Properties, states, and qualities of objects and processes related to education:
  - (5) *...relationship between critical thinking skills and academic performance...*
  - (6) *...cognitive abilities and domain-specific performance...*
- (4) People or individuals who participated in processes or operations in the field of education:
  - (7) *...students who experience test anxiety benefit from...*
  - (8) *Highly adaptive teachers possess a heightened diagnostic perception...*
- (5) Places or settings where educational actions or processes took place:
  - (9) *...student attainment in the early years of school...*
  - (10) *...the proportion of German speakers (L1) in a class was a predictor...*
- (6) Factors involved in or affecting educational processes:
  - (11) *...the negative emotions such as anger, anxiety...*
  - (12) *...intrinsic motivation may play an important role...*
  - (13) *...individual interest can fuel knowledge acquisition...*

The terms analyzed were related to all the other terms that formed part of the same subject field, educational psychology, constituting a conceptual system. The concepts that

these terms represented shared characteristics of this discipline and were connected to each other through logical and ontological relationships (Cabré, 1999) as shown in the diagram (Figure 1). As regards logical relationships, the concepts represented by these terms were related through logical subordination and coordination since they shared the characteristics of a general concept, but each of them also had at least one distinguishing feature. For example, the concepts designated by the words *learning*, *performance*, *thinking* and *understanding* represented cognitive processes which took place in educational contexts, but each process had its own characteristics; the concepts *students* and *teachers* both represented the people involved in those processes but each of them had a different role. Regarding ontological relationships, it is interesting to notice that chain or cause-effect relationships were established among the concepts designated by the terms *school*, *students*, *teachers*, *learning*, *performance*, *content*, *understanding*, *ability* and *interest*. The cause-effect relationships were evident in the fact that *students' ability* and *interest*, and their *understanding* of the *content*, which was mainly delivered by *teachers*, determined or influenced the *students' learning* process and *performance* at *school*. The combined logical and ontological relationships determined the hierarchical structure of the subject field of Educational Psychology, as shown in Figure 1.

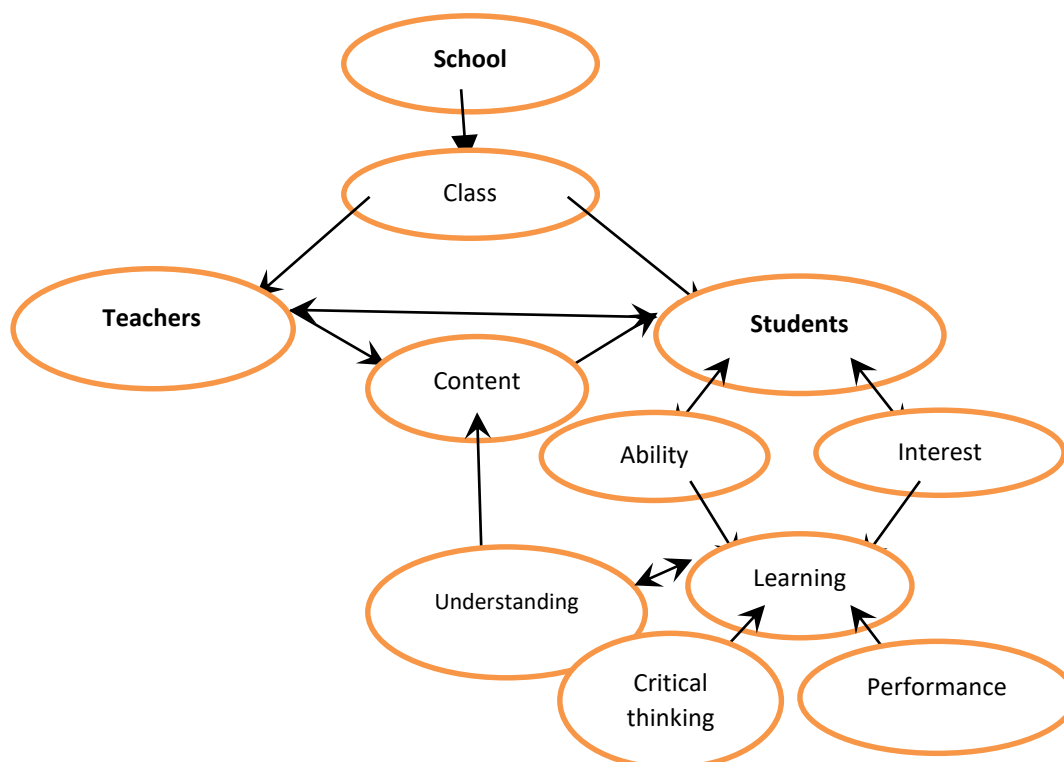


Figure 1. Diagram of the conceptual field of GSL-Lex words with technical meaning

The technicality of these words was also shown in their tendency to co-occur with other lexical items of the subject field (Adelstein & Kuguel, 2005; Cabre, 1999; Schmitt, 2000). The concordances of these items revealed that the combinations were mostly adjective + noun collocations, for example, *verbal tasks* or *verbal styles*, and noun + noun such as *outcome emotions* or *achievement emotions*. In all these examples, both words contributed to the meaning of the collocation and their combination made them technical. In the context of educational psychology, these recurrent combinations were terminological

phrases which signaled specific concepts in this field and revealed the collocational patterns typical of the specialized language used in this discipline, as argued by researchers in other disciplines (Gardner, 2007; Hyland, 2009; Hyland & Tse, 2007; Muñoz, 2015).

As regards the semantic behavior of these words, it is also important to note that some of these terms were words which are commonly used in other disciplines but with a different meaning. These items were redefined in the conceptual scheme of educational psychology and acquired specialized meanings in this field. An example is *performance* which, in the field of arts, refers to the act of performing a play, concert or some other form of entertainment, while in education it means the action or manner of carrying out a task with a particular learning goal.

Pragmatically, these words were regarded as technical as they were used in a specialized communicative situation. These field-specific words were used by educational psychologists to communicate specialized topics of the discipline in a specific genre, the research article.

### **Words with academic meaning**

Within the most frequent words of the corpus, more than half of the words coming from the AWL did have academic meaning, for example *significant, variables, response, specific, negative, previous*, but many words with academic meaning (79) came from the GSL-LexW, such as *model, also, suggests, test, respectively, important, current, likely, multiple*, and 21 words came from the List of Other words, such as *scores, correlations, questionnaire, predictor*. These words had academic meaning in our corpus since they referred to the linguistic acts expressed in the texts and to the relationships established among clauses and sentences by the writers and researchers in the articles analyzed. These words also represented concepts related to the methodology and research processes carried out in the field of educational psychology. These words were reclassified as academic following the categories proposed by Meyer (1990 in Coxhead & Nation, 2001) and Meyer (1997), as described in Section 2.2.3. Some examples are provided below:

#### **(1) The domain of the text:**

(a) Vocabulary related to linguistic acts performed in a text:

(14) *...research has also indicated that scale results can vary...*

(15) *We suggest that the performance differences are associated...*

(b) Textual deixis:

(16) *No differences emerged between women and men, respectively, for...*

#### **(2) Elements of scholarly practice:**

(a) Vocabulary referring to scientific activities:

(17) *All participants completed a one-page questionnaire based on...*

(18) *We used a multiple-choice test to assess ...*

(19) *Average age of the total sample was 16.0 (SD=1.24)*

(b) Properties of scientific activities:

(20) *...greater frequency of DA elevations were specific to males...*

(c) Evaluations of theories, procedures, methods, results:

(21) *...this potentially negative cycle of affect and interactions...*

(22) *...perceived complexity are important for predicting interest...*

(d) Classifiers of data:

(23) *...inclusion of children's response to learning-related challenges...*

(24) *...the treatment effects on the total posttest scores...*

#### **(3) Vocabulary referring to the subject matter of scientific activities**

(a) Lexical expressions referring to tense, aspect, and modality:

- Temporal deixis:
  - (25) *A strength of these previous person-centered studies is the use of...*
  - (26) *The current data are consistent with these prior results...*
- Modality:
  - (27) *...demonstrating that children are more likely to delay gratification...*
  - (28) *...it was not possible in this study to separate children...*
- Event quantifiers:
  - (29) *...to choose multiple correct answers among a number of alternatives...*
- (b) Abstract quantities or quantitative properties of states of affairs or entities:
  - (30) *...the total posttest scores...*
  - (31) *...low levels of delay of gratification...*
- (c) Classifiers of states of affairs:
  - (32) *...the augmented Cox regression model...*
  - (33) *... The strongest predictor of academic performance...*
- (d) Relations between states of affairs:
  - (34) *...incentives can also increase the degree of extrinsic motivation...*
  - (35) *...the correlations between playtime and study time were inconsistent...*
  - (36) *Static diagrams included explicit spatial information...*
- (4) Discourse-organizing vocabulary:**
  - (37) *Also, girls showed a smaller positive link...*
  - (38) *...therefore, gender was excluded from further analyses...*

## Discussion

Our vocabulary study, based on pragmatic and semantic criteria, allowed us to produce a vocabulary list specific to the genre and field analyzed. This analysis rendered a list which was made up of only academic and technical words related to the field of educational psychology. The classification showed that none of the high frequency words coming from the GSL had general meaning in our corpus, and not all the words from the AWL were used with academic meaning.

The results revealed that, of the 237 words selected as most frequent, around 60% of the items had academic meaning, but the rest were all technical words, which indicates that none of the words had general meaning, according to semantic and pragmatic criteria (Cabre, 1999; Meyer, 1997; Pearson, 1998). Even those items from the group of *Other words*, which could have been expected to be technical if considering Nation's (2001) criterion, were either academic or technical.

The technical words represented concepts which shared characteristics of the field of educational psychology and were related to each other through semantic relationships (Cabr e, 1999). Many of these words usually co-occurred with other words, forming terminological phrases with field-specific meanings, as observed by Hyland and Tse (2007). The academic words referred to processes or actions which a writer carries out in a text, and to concepts related to the methodology of the science, the stages of the research process and their evaluation (Meyer, 1990 as cited in Coxhead & Nation, 2001; Meyer, 1997). It is also important to highlight that, as reported by Cabre (1999), a large number of words with technical and academic meaning were terms of Latin origin, which may have a crucial impact on the comprehension of educational psychology research articles by readers whose mother tongue is a romance language.

It was possible to observe that, in our corpus, the usefulness of the traditional wordlists was undermined as, in relation to the GSL, none of the lexical words from the list were used as general words, and, in relation to the AWL, many of its words had technical meaning.

Specifically, the findings showed that there were many items in our corpus that are commonly used in the general language, but they acquire specialized technical meanings in the field, as observed by Sutarsyah, Nation and Kennedy (1994) and Chung and Nation (2003) in an economics text and an applied linguistics text, respectively. These results support Adelstein and Kuguel's (2005) conclusion that in the social sciences a large proportion of the technical words are general words which are redefined semantically in a specific conceptual field and may take on specialized meanings in different disciplinary environments. This highlights the notion of disciplinary specificity as argued by different authors such as Garofolo et al. (2009), Hyland (2000; 2009), Hyland and Tse (2007) and Wang et al. (2008).

Our findings support Hyland and Tse's (2007) observation that the vocabulary of the existing wordlists can be too general for students who need to read field-specific texts in academic contexts. They also add to the criticisms of the usefulness of the GSL and the AWL as wordlists of value to students of specific disciplines involved in academic studies (Hancioglu, et al. 2009; Ward, 2009). Our results lend support to the need to build specialized lists based on corpora representing specific genres and disciplines, as suggested in previous studies (Garofolo et al., 2009; Chen & Ge, 2007; Hyland, 2009; Hyland & Tse, 2007; Muñoz, 2015; Vongpumivitch et al., 2009; Wang et al., 2008). Further research is needed to confirm the usefulness of building specific wordlists from corpora, without using previous lists as baseline.

### **Pedagogical implications**

Our study highlights the relevance of creating small, specialized corpora with the texts that EAP learners need to read in specific disciplines, which can offer the instructors the possibility to identify, select and teach a manageable number of words which occur frequently in these texts. Our results may specifically address the needs of non-native psychology students at UNRC, by providing them with a description of the vocabulary of the educational psychology research article (EPRA), and a list of the most frequent academic and technical words representative of this field and this genre. This field-specific list may help learners study the EPRA vocabulary in a more conscious and manageable way. This list may provide a clear and direct access to the most frequently used words in this specific genre and guide students' attention to identifying and understanding the features of the discourse which they usually read in their particular courses.

The findings of our study may contribute to the psychology students' better understanding of the vocabulary of the EPRA and help them reach the lexical threshold which is necessary to comprehend this genre. The list of highly frequent academic and technical words provided in this study may not only facilitate our students' process of reading comprehension but also serve as a useful teaching resource for EAP instructors to carefully plan the curriculum, create reading materials and design relevant activities for vocabulary learning.

Considering that this study was carried out using a small corpus in a restricted field, an interesting area of further research would be to study this vocabulary in a larger corpus of educational psychology research articles to be able to expand our results and make generalizations for this field.

### **Acknowledgement**

This study was supported by a grant of SeCyT UNRC (Secretaría de Ciencia y Tecnología, Universidad Nacional de Río Cuarto, Argentina).

## References

- Adelstein, A., & Kuguel, I. (2005). La lectura y la escritura de textos que comunican ciencia: la experiencia del Taller de Lectoescritura del Primer Ciclo Universitario. In *II Jornadas sobre docencia. Los docentes universitarios ante los nuevos escenarios para la formación de los estudiantes* (pp. 59-66). Bs. As.: Universidad Nacional de General Sarmiento. Secretaría Académica.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (2000). Longman grammar of spoken and written English. Essex: Longman.
- Brezina, V. & Gablasova, D. (2013). Is There a Core General Vocabulary? Introducing the New General Service List. *Applied Linguistics*, 1(23), 1-23.
- Cabré, M. (1999). Terminology. Theory, methods and applications. Barcelona.
- Chen, Q., & Ge, G. (2007). A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles. *English for Specific Purposes*, 26, 502-514.
- Chung, M., & Nation, P. (2003). Technical vocabulary in specialised texts. *Reading in a foreign language*, 15(2), 103-116.
- Chung, M., & Nation, P. (2004). Identifying technical vocabulary. *System*, 32, 251-263.
- Coxhead, A. (2000). A New academic word list. *TESOL Quarterly*, 34 (2), 213-238.
- Coxhead, A. (2011). The Academic Word List ten years on: Research and teaching implications. *TESOL Quarterly*, 45(2): 355-362.
- Coxhead, A., & Nation, P. (2001). The specialised vocabulary of English for academic purposes. In J. Flowerdew & M. Peacock (Eds.). *Research Perspectives on English for Academic Purposes* (pp. 252-267). Cambridge: Cambridge University Press.
- Durrant, P. (2016). To what extent is the academic vocabulary list relevant to university student writing? *English for Specific Purposes* 43, 49-61. Elsevier. [www.journals.Elsevier.com](http://www.journals.Elsevier.com)
- Gardner, D. (2013). Exploring Vocabulary: Language in Action. Routledge.
- Gardner, D. & Davis, M. (2013). A New Academic Vocabulary List. *Applied Linguistics* 35(3), 305-327.
- Gardner, D. (2007). Validating the construct of *word* in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics* 28(2), 241-265.
- Halliday, M.A.K. (1985). *An introduction to functional grammar*. London: Arnold.
- Hancioglu, N., Neufeld, S. & Eldridge, J. (2008). Through the looking glass and into the land of lexico-grammar. *English for Specific Purposes* 27, 459-479.
- Hyland, K. (2000). *Disciplinary Discourses. Social interactions in academic writing* London: Pearson Education.
- Hyland, K. (2006). *English for academic purposes. An advanced resource book*. London: Routledge.
- Hyland, K. (2009). Writing in the disciplines: Research evidence for specificity. *Taiwan International ESP Journal*, 1(1), 5-22.
- Hyland, K., & Tse, P. (2007). Is there an "academic vocabulary"? *TESOL Quarterly*, 41(2), 235-253.
- Hyon, S. (1996). Genre in three traditions. *TESOL Quarterly*, 30 (4), 693-719.
- Levine, D., & Stephan, D. (2010). *Even you can learn statistics*. New Jersey: Pearson Education.
- Lei, L. & Liu, D. (2010). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes* 22, 42-53.
- Li, Y. & Qian, D. (2010). Profiling the Academic Word List (AWL) in a financial corpus. *System* 38, 402-11.

- Meyer, P. G. (1997). *Coming to know. Studies in the lexical semantics and pragmatics of academic English*. Tübingen: Narr.
- Muñoz, V. (2015). The vocabulary of agriculture semi-popularization articles in English: A corpus-based study. *English for Specific Purposes*, 39, 26–44.
- Nation, I. S. P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Pearson, J. (1998). *Terms in Context*. Amsterdam: John Benjamins Publishing Company.
- Schmitt, N. (2000) *Vocabulary in Language Teaching*, Cambridge: Cambridge University Press.
- Scott, M. (2004). *Wordsmith tools*. Oxford: Oxford University Press.
- Scott, M. & Tribble, C. (2006). *Textual patterns*. Amsterdam: John Benjamins.
- Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2005). *Corpus and text. Basic principles*. En M. Wynne (ed.) *Developing linguistic corpora: A guide to good practice*. Oxford: Oxbow Books: 1-16. Available online: <http://ahds.ac.uk/linguistic-corpora> [Accessed: December, 22<sup>nd</sup>, 2006.]
- Stubbs, M. (1986). *Educational Linguistics*. Oxford: Basil Blackwell.
- Sutarsyah, C., Nation, P., & Kennedy, G. (1994). How useful is EAP vocabulary for ESP? A corpus based study. *RELC Journal*. 25 (2), 34-50.
- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Glasgow: Cambridge University Press.
- Vongpumivitch, V., Huang, J., & Chang, Y. (2009). Frequency analysis of the words in the academic word list (AWL) and non-AWL content words in applied linguistics research papers. *English for Specific Purposes*, 28(1), 33-41.
- Wang, J., Liang, S., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27, 442-458.
- Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes* 28, 170–182.
- West, M. (1953). *A general service list of English words*. London: Longman.



### Appendix

List of frequent academic and technical words of the educational psychology research article  
Words with academic meaning (ranked according to frequency of words)

Word	Word	Word
1 test	48 value	95 therefore
2 et	49 number	96 total
3 al	50 whether	97 then
4 e (e.g./i.e.)	51 thus	98 view
5 group	52 further	99 point
6 significant	53 factor	100 provided
7 also	54 included	101 small
8 g (e.g.)	55 current	102 n (population of the data set)
9 items	56 given	103 influence
10 research	57 hypothesis	104 so
11 p (p value)	58 frequency	105 following
12 related	59 example	106 often
13 model	60 c (questionnaire for children)	107 validity
14 participants	61 possible	108 conducted
15 results	62 t (regression coefficient)	109 fig (figure)
16 high	63 correlations	110 revealed
17 used	64 addition	111 fit
18 scores	65 r (correlation coefficient)	112 response
19 variables	66 average	113 role
20 effect	67 concept	114 types
21 level	68 order	115 set
22 scale	69 overall	116 suggests
23 approach	70 greater	117 very
24 information	71 previous	118 expected
25 found	72 consistent	119 questionnaire
26 positive	73 d (dimension / deviation)	120 multiple
27 based	74 identified	121 prior
28 analysis	75 focused	122 required
29 specific	76 mean	123 future
30 low	77 cross	124 regression
31 see	78 compared	125 way
32 table	79 interaction	126 fact
33 perceived	80 b (unstandardized variable)	127 final
34 only	81 s (second)	128 need
35 measures	82 similar	129 range
36 sample	83 showed	130 contrast
37 goal	84 f (null hypothesis test)	131 section
38 reported	85 evidence	132 completed
39 important	86 examine	133 furthermore
40 data	87 terms	134 predicted
41 present	88 indicated	135 predictor
42 negative	89 observed	136 relative
43 m (mean)	90 same	137 taking
44 likely	91 due	138 theory
45 well	92 questions	139 respectively
46 SD (standard deviation)	93 support	
47 associated	94 rather	

## Words with technical meaning (ranked according to frequency of words)

<b>Word</b>	<b>Word</b>	<b>Word</b>
1 students	34 teachers	67 control
2 self	35 affect	68 mastery
3 performance	36 area	69 course
4 achievement	37 skills	70 males
5 learning	38 understanding	71 grade
6 time	39 verbal	72 new
7 academic	40 size	73 females
8 cognitive	41 avoidance	74 reasoning
9 differences	42 feedback	75 training
10 ability	43 solving	76 English
11 children	44 year	77 situation
12 motivation	45 regulation	78 state
13 task	46 mental	79 engagement
14 anxiety	47 experience	80 attention
15 reading	48 efficacy	81 visual
16 strategies	49 management	82 difficulty
17 mathematics	50 social	83 intrinsic
18 interest	51 comprehension	84 content
19 thinking	52 styles	85 effort
20 problems	53 classroom	86 nature
21 school	54 general	87 competence
22 emotions	55 beliefs	88 activities
23 knowledge	56 work	89 efficiency
24 class	57 language	90 personal
25 gender	58 memory	91 instructions
26 critical	59 outcomes	92 college
27 education	60 intelligence	93 development
28 spatial	61 domain	94 exam
29 cultural	62 processes	95 boys
30 age	63 orientation	96 quality
31 assessment	64 personality	97 success
32 science	65 instructional	98 settings
33 individual	66 behavioral	