Semantic Preference across Research Articles: A Corpus-Based Study of Adjectives in English

Valdênia Carvalho e Almeida

Abstract—The goal of the present study is to investigate the semantic preference of the most frequent adjectives in research articles through a corpus-based analysis of texts published in journals in Applied Linguistics (AL). The corpus used in this study contains texts published in the period from 2014 to 2018 in the three journals: Language Learning and Technology; English for Academic Purposes, and TESOL Quaterly, totaling more than one million words. A corpus-based analysis was carried out on the corpus to identify the most frequent adjectives that co-occurred in the three journals. By observing the concordance lines of the adjectives and analyzing the words they associated with, the semantic preferences of each adjective were determined. Later, the AL corpus analysis was compared to the investigation of the same adjectives in a corpus of Chemistry. This second part of the study aimed to identify possible differences and similarities between the two corpora in relation to the use of the adjectives in research articles from both areas. The results show that there are some preferences which seem to be closely related not only to the academic genre of the texts but also to the specific domain of the discipline and, to a lesser extent, to the context of research in each journal. This research illustrates a possible contribution of Corpus Linguistics to explore the concept of semantic preference in more detail, considering the complex nature of the phenomenon.

Keywords—Applied linguistics, corpus linguistics, chemistry, research article, semantic preference.

I. INTRODUCTION

THE main goal of the current corpus-based study is to determine the semantic preferences of high-frequent adjectives in research articles published in major journals in the field of AL. Additionally, the study compares the results from the AL corpus-analysis to the findings from the investigation of the same adjectives in a Chemistry corpus.

The concept of semantic preference comes from the notion that "many uses of words and phrases show a tendency to occur in a certain semantic environment." [3]. For example, the word (or lexical item) *large* frequently associates (or collocates) with words for "quantities and sizes", such as numbers, scale, part, amounts, quantities. The concept is also elaborated in [7], [2], and [6]. Reference [7] defines semantic preference as "the relation, not between individual words, but between a lemma or word form and a set of semantically related words". This set of semantically related words has traditionally been known as a lexical field (or semantic field): a group of words which share some semantic feature.

According to [2], as semantic preference is believed to be

Valdênia Carvalho e Almeida is with the Federal University of Viçosa, Viçosa, MG   36570-000 Brazil (phone: 55 31 98046-4880; e-mail: valdenia@ufv.br).

dependent on the register, context, and domain, it is probably shared among speakers of a given community. Hence, those who seek to become part of their academic community need to be familiar with the language patterns being used. These patterns include word choices, word associations, phrases, and others.

The description of semantic preference has been the focus of studies in Corpus Linguistics – CL [1], [5], [8]. The increasing use of CL methods has allowed researchers to identify systematically sets of words co-occurring in language. In other words, the relationship between an item and its environment may be best revealed thorough CL methods and tools. Concordance lines, for example, clearly reveal repetitions that we can analyze quantitative and qualitatively.

The present research has been founded on the belief that semantic preference is register- and domain-dependent, and thus, we consider necessary to investigate semantic preference in specific registers in order to find out the kind of language being used in different domains. Based on this assumption, research articles from different disciplines are expected to show some variation in language use. In the present paper, the semantic preferences of six adjectives will be examined, namely *different, high, important, new, same,* and *significant.*

II. METHODS

The specialized corpora used in the present research comprise text collected from journals in the fields of AL and Chemistry. The AL corpus contains research articles published in the period from 2014 to 2018 in the three following journals: English for Academic Purposes (EAP), Language Learning and Technology (LLT), and TESOL Quaterly (TQ), totaling 1,053,651 words. Each journal was considered as being a sub-corpus of the AL corpus, and the three sub-corpora were analyzed separately in the first part of the study. The Chemistry corpus, which was used only in the second part of the study for matters of comparison, comprises 1,053,317 words of text collected from journals in the field. The corpus of Chemistry also comprehends publications from 2014 to 2018.

In the first part of the research, the software *AntConc* 3.5.7 [4] was used to identify the most frequent adjectives that co-occurred in the three journals in AL. This paper presents the analysis of six of the most frequent adjectives identified in the sub-corpora, namely *different, high, important, new, same*, and *significant*. Each sub-corpus was analyzed separately for the frequency of occurrence of the adjectives to make sure to select adjectives that frequently co-occurred in the three journals. Table I shows the frequency of the adjectives in each

journal separately and their total frequency of occurrence in the AL corpus. The adjectives are presented in alphabetical order.

TABLE I
FREQUENCY OF THE ADJECTIVES ACROSS THE JOURNALS IN AL

Adjective	EAP (351,134 tokens)	LLT (351,162 tokens)	TQ (351,355 tokens)	Total (1053,651)
Different	721	552	606	1879
High	292	282	358	932
Important	288	203	345	836
New	237	344	267	848
Same	312	254	280	846
Significant	290	439	313	1042

In the three sub-corpora the six adjectives are among the most frequent ones, with some differences in the order of frequency of occurrence in each journal, as shown in Table II.

TABLE II
ORDER OF FREQUENCY OF THE ADJECTIVES IN EACH AL JOURNAL

EAP	LLT	TQ
1. different	1. different	1. different
2. same	2. significant	2. high
3. high	3. new	3. important
4. significant	4. high	4. significant
5. important	5. same	5. same
6. new	6. important	6. new

After selecting the most frequent items to be investigated, the nouns associated with each adjective, with a co-occurrence frequency of 3 or above, were identified and grouped into semantic sets. These sets were then analyzed in order to determine the semantic preferences of the adjectives within the three journals from AL. In this part of the analysis concordance lines of the adjectives were studied manually, making it possible to observe the meaning of the words in context and to place them into a lexical or semantic field. For instance, by examining the list of collocates and the concordance lines of the adjective *high* we could observe that some of the words that it is associated with were *expectation, anxiety, engagement , frustration, motivation*, and *satisfaction*. Hence, these words were classified under the same semantic field comprehending words related to *emotions, feelings*, and *behavior*, which, in turn, constituted one of the semantic preferences of the adjective *high* in the AL corpus.

It is important to clarify that the present analysis had its focus on the nouns associated with the six adjectives used in attributive position. That means the discussion presented here will sometimes mention some of the adjective collocates involved in the associations, but for matters of space, further details will not be counted for. The word *background* for example, associates with *different* forming the combinations *different cultural background, different educational background*, and *different background* in the AL corpus. The study counted these combinations as three occurrences of *background* with *different* without making any distinctions among them. An example from the Chemistry corpus would be the noun *property* associated with the adjective *different* in the combinations *different chemical properties, different physical properties*, and *different properties*. The analysis of

other lexical items used in the near environment of the six adjectives might be the focus of further investigation. Furthermore, as the study assumes that semantic preference is register- and domain-dependent, data analysis and interpretation was performed taking into account the meanings and communicative functions of the associations in relation to the academic register and research area. That is, semantic sets and semantic preferences were determined in relation to the specific context of research articles in specific academic fields. In that regard, the semantic sets and semantic preferences identified in each corpus (AL and Chemistry) were compared to verify how distinct the language in the two fields are in relation to the use of the six adjectives under study. For example, when comparing the use of *high* in the two corpora we noticed that in the AL corpus it is more commonly used with words related to quantification and measurement whereas in the corpus of Chemistry it associates with words from Chemistry domain. For example, the association *high level of* had 58 occurrences in the AL corpus against 15 occurrences in the corpus of Chemistry, even though the total number of occurrences of *high* in Chemistry was much higher. This comparative analysis was carried out as a second part of the study, and in addition to find out some differences between the language used in both areas, it tried to relate the communicative functions of the adjectives to the sections comprising research articles, namely introduction, methods, results, discussion, and conclusion.

The first step in this comparative analysis was to check the frequency of occurrence of the six adjectives in the corpus of Chemistry. Table III shows the total raw frequencies and the normalized frequencies (per thousand of words) of the lexical items in the chemistry and also in the AL corpus.

TABLE III
FREQUENCY OF OCCURRENCE OF THE ADJECTIVES IN AL AND CHEMISTRY CORPORA

Adjective	Frequency in Chemistry corpus (1053,317)		Frequency in AL corpus (1053,651)	
	Raw freq.	Norm. freq.	Raw freq.	Norm. freq.
Different	1867	1.78	1879	1.79
High	1648	1.57	932	0.88
Important	394	0.38	836	0.79
New	298	0.28	848	0.81
Same	978	0.93	846	0.80
Significant	495	0.47	1042	0.99

A glance at Table III shows that there were significant differences in the frequency of occurrence of the adjectives *high, important, new*, and *significant* in the two corpora. *High* was used almost twice times in the Chemistry corpus, whereas *important, new*, and *significant* occurred more than twice in the corpus of AL.

The analysis of the use of the adjectives in the field of Chemistry concerning their semantic preferences also revealed some interesting differences when compared to the use in AL. In the next part of the paper all the results concerning the semantic fields and semantic preferences identified in both corpora are presented and discussed, followed by the discussion concerned the comparative analysis.

III. RESULTS AND DISCUSSION

Taking the AL corpus first, all the words associated with the six adjectives were examined in context and grouped into semantic sets. Each semantic set was interpreted in terms of semantic preference. Tables IV-IX show the most common words or collocates that associated with the adjectives and their semantic preferences. In each table, the most common collocates are displayed in groups that correspond to their semantic sets, with the number of occurrences in the AL corpus in parentheses. The written form of each noun included in the tables corresponds to the most frequent form (singular or plural) of the noun used in the corpus. For example, if in the table the noun appears in singular form, it means it occurred more often in singular form in the corpus. If the noun is displayed in the table in its plural form it means it mostly appeared in the corpus in its plural form. As mentioned before, the focus of the analysis was on the noun collocates.

A. Semantic Preferences in AL Corpus

1. Different

TABLE IV
COLLOCATES AND SEMANTIC PREFERENCE OF *DIFFERENT* IN AL JOURNALS

Collocates	Semantic preferences
Contexts (38), backgrounds (22), disciplines (22), genres (19), languages (12), environments (9), meanings (9), texts (9), registers (8), discourse (3)	Words related to AL
Fields (15), patterns (14), functions (13), topics (11), areas (8), needs (8), cultures (7), technologies (3)	General research terms
Types of (87), ways (35), levels (30), aspects of (29), groups (26), forms (19), types (16), categories (15), parts of (14), kinds of (11), stages of (11), stages (10), factors (8), sections (0/8), classes (6)	Classification and partition
Strategies (37), conditions (34), resources (21), approaches (20), tasks (15), activities (12), tools (7), methods (5), tests (5)	Methods and procedures or tools of research
Perspectives (26), views (7), perceptions (4), understandings (4)	Knowledge and way of thinking
Levels of (45), measures (11), amounts of (8), degrees of (4)	Quantification and measurement

As it was expected, a first look at Table IV shows several words commonly used in academic texts such as, *approach, category,* and *method*, and words related to the specific field of language learning study, for example, *genre, discourse*, and *register*. This is only to confirm the assumption that semantic preference is register- and domain-dependent. It was based on this assumption that the context of use of each association adjective + collocate was analyzed in order to place each word into the most appropriate semantic field. For example, the noun *environment* fell into the set of words related to field of study and journal topic because it occurred in combinations such as *different linguistic environments, different semantic environments, different learning environments*, etc. An applied linguist would recognize *environment* as a commonly used word in language learning study domain.

The study of the data showed that *different* associated more frequently with words related to the field of AL, classification and partition, methods and procedures, knowledge and way of thinking, and quantification and measure. *Different* was mainly used in the corpus to express diversity and to make comparisons.

(1) I classified verbs with **different meanings** as they are used in **different contexts**.

(2) But there are **different registers** in written English (e.g. written academic prose, news and fiction) …

(3) Nevertheless, the Camtasia video suggests that the two groups had **different strategies** for their dictionary use.

(4) We say "could be interpreted" because we acknowledge that there are **different perspectives** in understanding and defining fallacies …

(5) Despite the **different amounts of** vocabulary gains between the CODI and CONC conditions (t = 3.41, p < 0.001) …

The analysis of the use of *different* throughout the research articles reveals that in the introduction and discussion sections the associations with *different* are more related to diversity. That means the associations indicate that there are various. In methods and results sections the associations usually refer to comparisons.

2. High

TABLE V
COLLOCATES AND SEMANTIC PREFERENCE OF *HIGH* IN AL JOURNALS

Collocates	Semantic preferences
Level of (58), scores (27), rate (22), -scoring (19), degree of (14), proportion (14), frequency of (12), correlation (10), frequency (8), numbers of (8), percentage (7), amount of (3), load (3)	Quantification and statistics
Proficiency (15), competence (6), experience (5), reliability (5), achievement (4),	Ability and quality
Expectations (11), anxiety (9), engagement (5), frustration (4), motivation (3), satisfaction (3)	Emotions, feelings, and behavior

The analysis of the concordance lines indicates that *high* was mainly associated with quantitative analysis and statistics in AL. It was also associated with words that illustrate ability (e.g. *high fluency*) and quality (e.g. *high importance*), and with words that express feelings (e.g. *anxiety*) and behavior (e.g. *engagement*).

The associations with *high* were mainly found in results and discussion sections.

(6) Only a few papers had a **high degree of** text matching.

(7) There are often **high expectations** for student writing at postgraduate level…

(8) …11participants were found to have **high experience.**

(9) The **high correlation** between AVST and XK-Lex scores suggests these two tests may well be fulfilling the same function.

The study of the lines also showed that the collocations *high degree of* and *high level of* had patterns of associations similar to the adjective *high* being used alone, especially when combined with words expressing ability (e.g. *high degree of proficiency, high level of fluency*) and emotions (e.g. *high degree of anxiety, high level of frustration*).

It is important to mention that, although not listed in Table V, a word that frequently collocated with *high* was *school*. The combination *high school* had 179 occurrences in the corpus, being 101 occurrences of *high school* or *high-school* functioning as adjective (e.g. *high school group, high school examinations, high school teachers*), and 79 occurrences of the noun *high school* indicating place (e.g. *academic preparation in high school, local high schools*). *High* was also associated with other lexical items forming compound adjectives such as, *high-level, high-quality, high-frequency, high-stakes, high-prestige, high-variability*, totaling 54 occurrences of *high-*.

3. Important

TABLE VI
COLLOCATES AND SEMANTIC PREFERENCE OF *IMPORTANT* IN AL JOURNALS

Collocates	Semantic preferences
Factor (27), issues (13), concept (8), function (8), questions (8), area (7), features (6), topic (3)	General research terms
Implications (17), findings (14), contribution (8), insights (8), consideration (7), evidence, caveat (3), result (3)	Words related to the structure of inquiry
Part of (15), aspects of (13), component (9), type of (6), means (5), aspect (4), criterion (3)	Classification and partition
Contexts (6), language resources (6), skills (6), practices (5), feedback (4)	Words related to AL

The adjective *important* was frequently associated with general research terms and with words related to the language learning study domain.

(10) This study raises a number of **important issues**, such as the role of teacher education in preparing and supporting teachers.

(11) …generally all understood the **important function** that such markers play in making text "considerate and accessible.

Interestingly, *important* is the adjective that was most associated with research terms closely related to the structure of research articles. Combinations such as *important implications, important findings,* and *important contribution* seem to be used in the corpus for the purpose of valuing research central issues.

(12) This study contributed **important findings** to the literature about the factors […]

(13) First of all, an **important contribution** of the present study was that it provided significant information about the use of metacognitive strategies […]

(14) In addition, the study **provides important insights** into some non-technological constructs.

4. New

Table VII illustrates the high diversity of words related to the field of AL that are associated with the adjective *new*. These associations were used to indicate innovation (*new technologies*), to refer to something that is different from one that existed earlier (*new practices, new roles*), and specially to refer to something not yet familiar (*new words, new vocabulary, new knowledge*). The use of *new* to refer to

something not yet familiar was frequently found in the introduction of the articles when contextualizing the study.

(15) …sensitivity to the norms of the target discourse community can open more possibilities to succeed in a **new community.**

(16) They assist in scientific thinking, and they help students learn **new knowledge** or transform their everyday knowledge…

(17) Collecting and analyzing CIs can help teachers to gain **new insights** about themselves.

(18) …readers need both traditional reading strategies and **new skills** associated with reading online texts.

TABLE VII
COLLOCATES AND SEMANTIC PREFERENCE OF *NEW* IN AL JOURNALS

Collocates	Semantic preferences
Words (37), literacies (34), vocabulary (28), practices (15), context (13), language (13), academic community (12), environment (10), skills (9), students (8), identity (7), pedagogies (3), phrases (3)	Words related to AL
Technologies (34), items (10), studies, (8), research (6), roles (6), questions (5), content (4), framework (4), themes (4), theories (4), features (3)	General research terms
Knowledge (28), ideas (15), information (10), Insights (8), perspectives (8), understandings (5), concepts (4), belief (2/1), paradigm (3),	Knowledge and way of thinking
Approach (15), method (10), strategy (8), tools (8), material (7)	Methods and procedures or tools of research

5. Same

TABLE VIII
COLLOCATES AND SEMANTIC PREFERENCE OF *SAME* IN AL JOURNALS

Collocates	Semantic preferences
Meaning (5/3), context (7), structure (7), word (7), book (6), genre (6), sentence (6), text (6), function (5), background (4), class (4), essay (4), language (4), token (4), verbs (4), vocabulary (4), article (3), content (3/0), discourse (3), form (3), register (3)	Words related to AL
Topic (14), items (11), pattern (8), results (8), features (6), study (4), data (3), findings (3), subject (3)	General research terms
Task (15), test (13), procedure (11), instruction (6), assessment (5), materials (5), approach (4), format (4), criteria (3), condition (3), tool (3),	Methods and procedures or tools of research
Course (14), student (11), group (8), teacher (8), university (8), program (6), person (5), school (5), interviewer (4), author (3)	People and location
Level (10), number of (10), level of (9), amount of (8), score (3), rate (3)	Quantification and measurement

By observing the collocates of *same* it can be seen that the adjective usually associates with words used in the description of the research methods and results. Associations such as, *same proficiency level, same school,* were frequently found in methods sections when presenting and describing participants and people or places related to the focus of research. When describing procedures or methods we observed constructions

such as *same approach, same procedure, same task, same test*, and the like.

(19) The purpose of the current study is to compare perceptions of international students' speech between two groups of instructors at the **same university**.

(20) The students were all assumed to be advanced at the **same level of** L2 proficiency.

(21) The **same procedure** was followed for the LREs that occurred during task performance for Research Question 2.

(22) All the groups performed the **same tasks**. During the tasks, the researcher orally provided the appropriate feedback type.

When presenting and describing research results s*ame* occurred in the combinations *same patterns, same results, same score,* and so on. In addition, the analysis of the concordances revealed that *same findings, same results, same study* were often used in the texts to compare the study to previous ones. Such comparison is a common practice in research as any study seeks to contextualize its findings within the larger body of research. The use of associations with *same* in this context can be an indication of possible generalization of findings or confirmation of effectiveness of certain methods or approaches. The samples below show some of the associations of *same* identified in the AL corpus.

(23) The native speakers showed the **same pattern** of genre effects as the L2 writers with regard to increased length-of-unit complexity.

(24) The authors reported the **same research findings** as the present study.

Moreover, a frequent use of the adjective in the research articles was as part of the lexical bundle *at the same time*, totaling 114 occurrences of *same* in the corpus.

6. Significant

TABLE IX
COLLOCATES AND SEMANTIC PREFERENCE OF *SIGNIFICANT* IN AL JOURNALS

Collocates	Semantic preferences
Difference (262), effect (117), correlations (54), main effect (44), interaction (42), predictors (40), relationships (17), levels (8), results (14), variation (4)	Statistics, quantitative analysis
Improvement (55), gains (23), increase (12), changes (11), decrease (3), enhancement (3)	A change or a process of change
Factor (10), implications (5), contributions (3)	Research general terms

The adjective *significant* was mainly used to refer to quantitative analysis and results (e.g. *significant correlations, significant main effect, significant improvement*). The association *significant difference* was the most frequent one and it was often used to report the results of quantitative comparisons.

(25) There was a **significant difference** in the post-test scores for text-only (M = 25.95, SD = 12.36).

(26) …there is a **significant correlation** between AVST and XK-Lex scores, i.e., if the two tests are measuring the same knowledge…

(27) …the ANOVA found a **significant main effect** of

discipline, F(1,80) = 6.608, p = 0.012.

Similarly, *significant* associated with words expressing a change or a process of change, such as *improvement, gain, change*, and *increase,* to refer to quantitative analysis and results.

(28) Results indicated **significant improvement** for the AV and A-only groups…

(29) …the experimental group evidenced **significant increases** in their intensity-ratios for intervocalic stops...

Also, *significant* associated with general research terms to emphasize products and claims of scientific research (e.g. *significant contributions, significant implications*). These combinations were more commonly found in discussion and conclusion sections.

(30) These frequency measures and the other available lexical measures have made **significant contributions** to lexical studies.

(31) The current study not only adds a more nuanced understanding of the interpersonal dimension of RA Discussions, but also has **significant pedagogical implications**…

Additionally, during the analysis of the concordance lines it was common to find the associations with the adjective *significant* preceded by the adverb *statistically*.

(32) The training period resulted in a **statistically significant increase** in the frequency of the target formulaic sequences

(33) …we found that there is **a statistically significant relationship** between these two measures

B. Comparative Analysis between AL and Chemistry Corpora

The second part of the study was dedicated to a comparative analysis between the corpus of AL and the corpus of Chemistry aimed to find out differences and/or similarities concerning the use of the adjectives in the texts published in the journals from both fields. To achieve this aim, we identified the semantic sets of the words associated with the six adjectives in the Chemistry corpus. The results for each adjective are shown in Tables X-XV.

1. Different

TABLE X
COLLOCATES AND SEMANTIC PREFERENCES OF *DIFFERENT* IN CHEMISTRY JOURNALS

Collocates	Semantic Preferences
Concentrations (52), temperatures (46), species (19), structures (19), mechanisms (15), materials (14), compositions (12), solvents (12), energies (11), selectivity (11), catalysts (10), environments (10), compounds (9), properties (9), components (8), molecules (7), reactions (6), cells (5)	Words related to Chemistry study domain
Conditions (36), models (20), behavior (17), methods (16), techniques (10), approaches (9), parameters (8), steps (8), strategies (5), procedures (4)	Methods, procedures, and results
Sizes (30), values (22), lengths (19), ratios (14), levels (13), amounts of (10), levels of (7), rates (7), loadings (6), coefficients (4), depths (3), degrees of (3), weights (3)	Quantification and measurement

In both corpora the adjective *different* was mainly used with words related to the domain of each discipline. Examples of some of these combinations are *different molecules* and *different reactions* in Chemistry, and *different genres* and *different registers* in AL. In addition to that, *different* was used with words that refer to the tools of research and procedures. Associations between *different* and the words: *conditions, approaches,* and *strategies* were found in the research articles from both areas, especially in the methods section.

By observing the other words used to refer to the tools of research it seems that each discipline has some preferences for some words against others.

Different also has a semantic preference for words that refer to quantification and measurement. Words comprising this semantic field, such as *sizes, values, ratios,* and so on are more common in the Chemistry corpus. A possible explanation is that quantitative analysis is more common in Chemistry than in Applied Linguistic research.

To sum up, there are some similarities concerning the semantic preferences of the adjective *different* in Chemistry and AL. In both corpora *different* is used with words related to the domain of each discipline, to the tools of research, and to quantification and measurement. However, the collocates of *different* in each corpus vary in number of occurrences and type. This variation may be related to the nature of research (quantitative, qualitative, and theoretical) usually conducted in each area.

TABLE XI
COLLOCATES OF *HIGH* IN CHEMISTRY JOURNALS

Collocates	Semantic Preferences
Temperature (199), activity (48), concentration (82) pressure (63), selectivity (40) surface area (37), density (46), content (39), stability (27), efficiency (23), sensitivity (22), purity (20), capacity (12), reactivity (8), solubility (8), conversion (7), productivity (4), magnetic field (3), porosity (3), precision (3)	Words related to Chemistry study domain
Values (17), level of (15), rates (13), number of (10), level (9), dispersion (8), amount of (8),	Quantification and measurement

The adjective *high* was used in the two corpora to refer to quantification and measurement, although this use is much more frequent in the Chemistry corpus. Again, this may be related to the nature of research, since quantitative analysis is much more common in Chemistry than in AL. The associations with *high* were more frequently found in the results and discussion sections.

It is interesting to note that *high* had almost twice occurrences in the Chemistry corpus than in the AL one. Several of these occurrences are not represented in Table XI because they refer to associations of *high* with other words forming compound adjectives. There were 128 occurrences of high in compounds. Some examples are: *high-accuracy, high-speed, high-resolution, high-energy, high-throughput, high-intensity, high-density, high-purity, high-sensitivity, high-pressure, high-temperature.*

2. Important

TABLE XII
COLLOCATES OF *IMPORTANT* IN CHEMISTRY JOURNALS

Collocates	Semantic Preference
Factor (22), parameter (15), step (5), aspect (4), implications (4), issue (4), result (4), topic (3), component (3), point (3)	General research terms

Important had twice occurrences in the corpus of AL (with a normalized frequency of 0.79) than in the corpus of Chemistry (normalized frequency of 0.38). Most of the associations with *important* in the Chemistry corpus was with general research terms. In both corpora the adjective was often found in results and discussion sections.

It is worth mentioning that the most-frequent word associated with *important* in both corpora was the noun *role*. There were 59 occurrences of role in the corpus of Chemistry in the following constructions: *play an important role (55), have an important role (2), identify the important role (1), the important role of (1).* In the AL corpus there were 39 occurrences of role with *important* in the constructions: *play an important role (33), the important role of (4), the important role they play (2), occupy an important role (1).*

3. New

TABLE XIII
COLLOCATES OF *NEW* IN CHEMISTRY JOURNALS

Collocates	Semantic Preferences
Method (13), phases (13), technique (7), data (6), insights (6), approach (4), methodology (3)	General research terms
Substance (9), compound (5), material (15), peak (14), feature (7)	Words related to Chemistry study domain

The adjective *new* occurred more than twice times in the AL corpus (normalized frequency of 0.81) than in the Chemistry one (normalized frequency of 0.28). In the corpus of Chemistry *new* was especially used with general research terms, and with words related to methods and procedures (*new methods, new technique, new approach,* and so on). The associations with *new* were often found in the methods and results sections. In the AL research articles, it seems that *new* associated with general research terms and words related to the field to indicate the need of new ways of learning a language (*new practices, new skills*), and to refer to something not yet familiar to the learner (*new vocabulary, new words*).

4. Same

The comparative analysis of the data revealed that in the research articles in both corpora the adjective *same* mainly occurs in the methods section to describe procedures of the inquiry. However, in Chemistry research articles *same* occurred more frequently in the results and discussion sections. That is, *same* was used not only to describe the procedures and conditions of the experiments conducted in the research but mainly to set out and discuss the study findings. In AL research articles *same* usually occurred in methods sections to describe the places and participants of the study (e.g. *same groups, same school*), and also to refer to research procedures (e.g. *same tasks, same instructions*).

TABLE XIV
COLLOCATES OF *SAME* IN CHEMISTRY JOURNALS

Collocates	Semantic Preferences
Temperature (27), concentration (12), dependence (9), phase (8), catalyst (7), features (6), morphology (3), property (5), structure (5), energy (4), atmosphere (3), atom (3), density (3), protein (3), peak (3)	Words related to Chemistry study domain
Condition (35), trend (24), sample (25), procedure (19), group (9), model (9), method (8), process (7), order of (12), order (8), pattern (6), setup (6), parameters (5), protocol (5), treatment (5), behavior (6), analysis (3), experiment (3)	Methods, procedures, and results
Amount of (10), size (9), level (8), value (8), loading (4), number of (4), volume of (4), level of (3)	Quantification and measurement

5. Significant

TABLE XV
COLLOCATES OF *SIGNIFICANT* IN CHEMISTRY JOURNALS

Collocates	Semantic Preference
Difference (39), change (31), increase (23), decrease (20), effect (19), loss of (12), reduction (12), improvement (11), amount of (10), deviation (7), correlation (4), enhancement (4)	Statistics, quantitative analysis

The adjective *significant* had more than twice occurrences in the corpus of AL (normalized frequency of 0.99) than in the corpus of Chemistry (normalized frequency of 0.47). In both corpora it was mainly associated with words related to quantitative analysis (*significant deviation*). However, in the AL corpus the associations with *significant* are sometimes less strictly related to quantitative analysis, especially when *significant* collocates with general research terms (e.g. *significant implication, significant contribution* and the like).

IV. CONCLUSION

The analysis of the collocates of the adjectives *different, high, important, new, same,* and *significant* in terms of semantic fields and semantic preferences has provided information that can contribute to the understanding of how language may vary across academic texts from different disciplines. Most of the semantic preferences identified in the AL corpus could also be identified in the Chemistry corpus. For instance, the adjective *same* showed semantic preferences for quantification (*amount of*) and methods/tools of research (*procedure*) in both corpora. However, many of the words comprising each semantic field varied across the two corpora. For example, the words related to methods used in each corpus in association with the adjective *same* were very different. In Chemistry *same* associated with *condition, protocol, model, treatment*, whereas in AL it associated with *approach, assessment, test, form*, and so on. As it can be seen, the two areas use different lexical items in association with *same* to refer to their methods and tools of research. Thus, the results indicate that depending on the discipline or the kind of research (quantitative, qualitative, and theoretical), there are some differences that may go beyond specialized vocabulary distinction. These include the word choices made by different academic communities when constructing the textual structure of the text, in terms of goals, methods and materials, results, discussion and conclusions. In this sense, it is important to continuously study the linguistic variation across registers, whether a macroscopic analysis (the study of the characteristics of whole texts) or a microscopic one (the study of a particular linguistic feature) is carried out. The detailed information provided by microscopic analysis can contribute to the teaching of EAP by better informing material developers and teachers about the specific linguistic features that learners should be aware of. Therefore, this study argues that second language teaching in the context of EAP should consider linguistic variation across disciplines. In the case of research articles, in addition to the distinction concerning the specialized vocabulary from different areas, there may be some differences in relation to the use of more general lexical items that should be considered. Each field or discipline might have different preferences when constructing the meaning of the scientific text and the structure of the inquiry. The current study can contribute for a better understanding of the kind of language that will fulfill the needs of those who seek to become part of their academic community.

REFERENCES

[1] A. Paciorek and J. N. Williams, "Implicit learning of semantic preferences of verbs," *Studies in Second Language Acquisition* , vol. 37, pp. 359 – 382, 2015.
[2] A. Partington, "Utterly content in each other's company: Semantic prosody and semantic preference," *International Journal of Corpus Linguistics* vol. 9, pp. 131–156, 2004.
[3] J. M. Sinclair, *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991, ch.1.
[4] L. Anthony, AntConc (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University, 2018. Available from http://www.laurenceanthony.net/software
[5] M. Begagić, "Semantic preference and semantic prosody of the collocation *make sense*," *Jezicolovlje*, vol. 14, pp. 403-416, Oct. 2013.
[6] M. Hoey, *Lexical Priming: A New Theory of Words and Language*. London - New York: Routledge, 2005, ch.2.
[7] M. Stubbs, *Words and Phrases: Corpus Studies of Lexical Semantics*. Oxford: Blackwell, 2001, ch.3.
[8] U. Oster and H. van Lawick, "Semantic preference and semantic prosody: A corpus-based analysis of translation-relevant aspects of the meaning of phraseological units," *Translation and Meaning*, part 8, pp. 333 – 344, Jan. 2008.

Valdênia Carvalho e Almeida. Brazil, 06/30/1975. Graduated in Language and Literature from Federal University of Viçosa (2002), master in Applied Linguistics from the Federal University of Minas Gerais (2007), doctor in Linguistic Studies from the Federal University of Minas Gerais (2014), and currently conducting her post-doctorate studies on linguistic variation at Sao Paulo Catholic University.

She has worked as a pedagogical coordinator and as an English teacher for many years. She has taught in Language Institutes, high-schools, and universities. Since the year of 2013, she has worked at the federal University of Viçosa as an assistant professor. In her current job, she teaches English for high-school students at the University of Viçosa and she also teaches Introduction to Corpus Linguistics and Using corpora in language teaching and learning for undergrad and graduate students. Her last publications have been in the use of Corpus Linguistics methods to investigate English learners writing namely "Result linking adverbials in learner corpora", *Domínios De Lingu@gem*, *13*(1), 400-431, and "The use of *make* in the writing of Brazilian English learners," *Domínios De Lingu@gem*, *8*(1), 75-90. Her current research interest is on the analysis of linguistics variation across different registers.

Dr. Almeida is a member of GELC, a research group that conducts researches in Corpus Linguistics in Brazil.