The Journal of Teaching Language Skills (JTLS)
6 (3), Fall 2014, Ser. 76/4
ISSN: 2008-8191. pp. 51-73

Key Lexical Chunks in Article Abstracts of 30 Applied Linguistics Journals

H. Farjami
Assistant Professor, TEFL
Semnan University
email: hfarjami@semnan.ac.ir

Abstract

In any discourse domain, certain chunks are particularly frequent and deserve attention by the novice to be initiated and by the expert to maintain a sense of community. To make a relevant contribution to the awareness about applied linguistics texts and discourse, this study attempted to develop lists of lexical chunks frequently used in the abstracts of applied linguistics journals. The abstracts from all the issues of 30 applied linguistics journals which were published before August 1, 2013 were collected. These abstracts which generated a corpus of 2,750,000 words were submitted to the program AntConc for chunk extraction. The long list of chunks in the output was shortlisted based on frequency and inclusiveness of shorter chunks. These were classified into textual and content n-grams. The article also presents the frequent chunks which serve as starting points in bringing up different aspects of research reports. The practical value of the results is briefly discussed at the end of the article.

Keywords: applied linguistics, article abstracts, lexical chunks, n-grams

1. Introduction

John Swales' (1990) seminal work on genre analysis has prompted interest in the analysis and *description* of academic journal articles, which represent the discourse of the academic community. Two primary lines of research have been pursued. Some studies have aimed at analyzing the macrostructures and move development in the major sections in research articles (e.g., Del Saz, 2011; Lim, 2006; Yang and Allison, 2003). The other

Received: 04/28/2014 Accepted: 10/15/2014

line has concerned itself with such micro-linguistic features of particular genres and text types as hedging (e.g., Malaskova, 2012), voice (e.g., Hinkel, 2004), tense (Hawes & Thomas, 1997), and reporting verbs (Bloch, 2010). As a major section in research articles, abstracts have received much attention both for their micwro- and macro structures. They deserve attention because they attempt to capture the essence of the whole article and are usually the first section in articles readers read (Hartley, 2003). However, one aspect of abstracts which seems to deserve special attention, but has been neglected to some extent, is chunks and collocations. The concern of this study is to add to the emerging body of facts and information about lexical chunks and collocative items which frequently recur in the abstracts of applied linguistics research articles. Such an attempt at identifying frequently occurring chunks can be of practical and functional significance, a point supported by Nation (2001), who maintains that frequent items usually have more utility and deserve more attention, particularly in educational domains.

2. Literature Review

2.1 Studies of abstracts

There are few scientific journals now which do not require an author to submit an abstract with the main research report. According to Ventola (1994), article abstracts are "tools of mastering and managing the ever increasing information flow in the scientific community" (p. 333). This means that research article abstracts play a key role in the academic and scientific sphere. This fact, coupled with Swales' (1990) promotion of ideas about the textual reality of genres, has encouraged some authors to study the structure and features of article abstracts as well as their variations across disciplines (e.g., Huckin, 2001; Hyland, 2004; Martin, 2003; Samraj, 2005). It has been established that abstracts differ from the main body of articles in their lexical, thematic and rhetorical structure and constitute a genre in their own right.

Among many academic disciplines, the abstracts of the articles in the field of applied linguistics, which are targeted by this study, have received some attention (e.g., Hyland, 2004; Lorés, 2004; Pho, 2008; Santos, 1996; Tseng, 2011). Hyland's (2004) study compared the move structure of abstracts across eight disciplines, including applied linguistics. Santos (1996) focused exclusively on the field of applied linguistics, examined 94 abstracts in applied linguistics articles, and found a prevalent five-move

model with sub-moves: 1. situating the research, 2. presenting the research, 3. describing the method, 4. summarizing the results, and, 5. discussing the results. Santos also examined the distribution of a few linguistic features such as verb tenses across moves. Lorés' (2004) small-scale study focused on the thematic organization of applied linguistics research article abstracts. Based on 30 abstracts from three journals, Pho (2008) explored the rhetorical moves of abstracts in the field of applied linguistics and educational technology as well as the linguistic realizations of the moves and the authorial stance in different moves in the abstracts. Finally, Farjami (2013) developed a corpus-based profile of the lexical make-up of applied linguistics research article abstracts and compared several categories of the most frequent applied linguistics words with those in two established wordlists, the Academic Word List and the General Service List, identifying the shared and unique items.

2.2 Chunks and collocations

Research in cognitive linguistics and psycholinguistics suggests that the basic units of language are constructions—pieces of language which are conventionalized in the community and represented in the minds of speakers and learners as language knowledge (Croft & Cruise, 2004; Goldberg, 2006). These stretches of language are similar to what Halliday (1966) called "collocations". The linguistic process involved is also reflected in the idiom principle introduced by Sinclair (1991, 2004). He articulated the principle as follows: "a language user has available to him or her a large number of semi-pre-constructed phrases that constitute single choices, even though they might appear to be analyzable into segments" (Sinclair, 1991, p. 110).

Unlike traditional approaches to language, which focus on the centrality of syntax, research has now established that lexical chunks and formulaic language are fundamental to the way language is used, processed, and acquired in both the L1 and L2 (Martinez & Schmitt, 2012). Martinez and Schmitt (2012), having cited research-based evidence for the essentialness of formulaic language, point out some of its important features: formulaic language is ubiquitous in language use; meanings and functions are often realized by formulaic language; formulaic language has processing advantages; formulaic language can improve the overall impression of L2 learners' language production (Martinez & Schmitt, 2012, pp. 300-301). Lewis (1997, 2000) took a more explicitly pedagogical interest in these

institutionalized expressions by arguing for a lexical approach to language teaching and stressing the inclusion and rehearsal in language programs of pre-patterned lexico-grammatical strings of words, which he called “lexical chunks”.

Whether pedagogically motivated or for descriptive purposes, some scholars have fruitfully compiled lists of formulaic sequences of words, most notably Shin and Nation (2008), Simpson-Vlach and Ellis (2010), and Martinez and Schmitt (2012). Shin and Nation (2008) tried to identify the most frequent collocations in spoken English. Simpson-Vlach and Ellis (2010), based on a wide range of academic genres, created a list of formulaic sequences for academic speech and writing and called it the Academic Formulas List (AFL), comparable to the Academic Word List by Coxhead (2000). Through a mixed-method corpus analysis, Martinez and Schmitt (2012) created a list of multiword lexical items including 505 phrasal expressions, whose frequency and pedagogical usefulness, they claim, make their list similar to well-established wordlists such as the GSL and AWL, and can be used in preparing tests and designing syllabi.

Research by those interested in English for Academic Purposes (EAP) has demonstrated that there is vocabulary and lexical chunks which are characteristically, if not uniquely, used in particular academic fields (Biber & Barbieri, 2006; Flowerdew & Peacock 2001; Hyland, 2004, 2008). However, most attempts at compiling lists of phrasal chunks or collocations have not been genre-specific and detailed. For example, although Simpson-Vlach and Ellis (2010) separately listed formulas that were specific to academic written language and academic spoken language and classified them according to their predominant pragmatic functions, they did not specify categories of chunks according to discourse communities or text types. This means that specific fields of discourse deserve further investigation and scrutiny in respect to lexical chunks or clusters.

3. The Study

The aim of this research was to explore and list the highest-frequency lexical chunks in applied linguistics article abstracts (ALAAs), the decision about the number of the frequent chunks to display in the report being made based on article space constraints and other practical limitations. More specifically, the study intended to report on two types of lexical chunks: chunks with a content orientation, and chunks with a textual or meta-discursive orientation. Along with this purpose, the following questions were formulated:

1. What are the most frequent content chunks in the abstracts of applied linguistics research journal articles?
2. What are the most frequent meta-discursive chunks in the abstract of applied linguistics research journal articles?

According to Hyland & Tse (2004), meta-discourse is a cover term which refers to "a range of devices writers use to explicitly organize their texts, engage readers, and signal their attitudes to both their material and their audience" (p. 156). However, it should be acknowledged that the textual expressions reported below may not cover the whole range of these functions in a fine-tuned and systematic way. There is also more to content chunks than what is outlined by this study. The modest and rough description in this study can be only a useful first step in shedding light on ALAA phraseology.

It might also be useful to note that many expressions are used in the literature to refer to the recurrent sequences of words which tend to appear together, including multi-word units, collocations, lexical chunks, lexical clusters, lexical phrases, n-grams, lexical bundles, idiomatic phrases, set phrases, prefabricated patterns, holophrases, formulas, lexicalized items, and phrasemes. N-grams, clusters, and chunks seem the most fitting terms to describe the language bites that were targeted in this study as they do not imply syntactic units and exclusively refer to pieces of language which are merely in immediate adjacency.

4. Methodology

4.1 The sample of applied linguistics journals

In the absence of established lines of demarcation for the field of applied linguistics (see for example, Davis, 2007; James, 2007), journal selection was entrusted to the judgment of the researcher, who was experienced in the field and familiar with its scope. The two lists of Humanities and Social Science in the Social Science Citation Index by Thomson Reuters were consulted and the journals which dealt with the field of applied linguistics were shortlisted. Out of this shorter list, the researcher then selected 30 journals which were better-known for publishing articles in applied linguistics and its subfield, e.g., language learning, language testing, CALL, SLA, and discourse studies. It should be admitted that a narrow sense of applied linguistics was born in mind and the main thrust in journal selection was language learning and teaching with some partiality toward

foreign/second language learning/teaching. (See Appendix for the list of the journals).

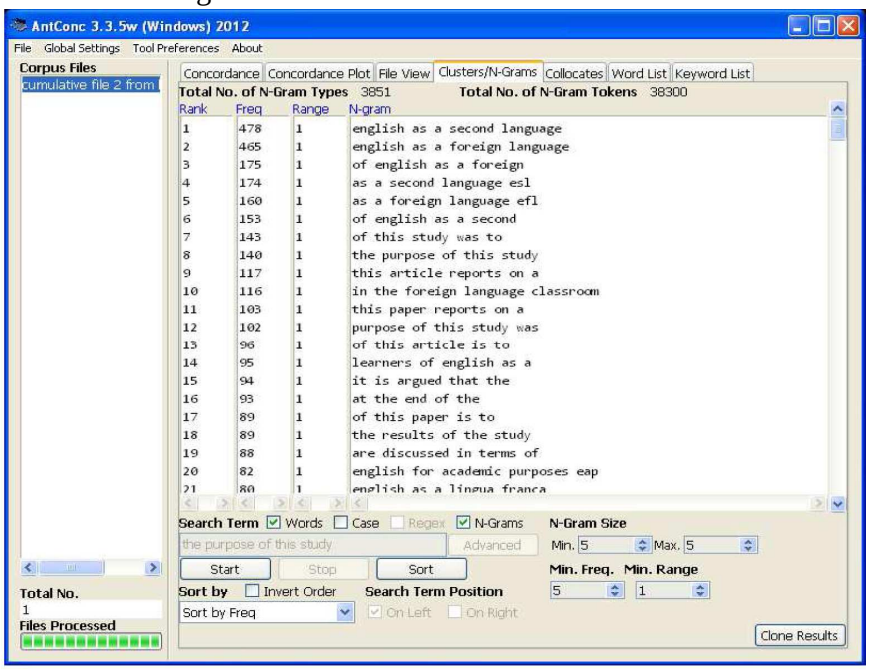
4.2 Preparing the corpus

This research is based on a corpus of 2,750,000 words coming from *all* electronically copyable abstracts in 30 major applied linguistics journals. The abstracts of feature articles in these journals were copied from their websites onto Word files. Then, information other than the titles and the main bodies of the abstracts was removed. The remaining texts were carefully reviewed for misspellings. Then, the files were converted to basic text format to take less computer memory space and be compatible with the text analysis software. Finally, the 30 files were coalesced in one massed file of 2,750,000 words. The dates of the publication of the abstracts ranged from the earliest issue of each journal which made free online abstracts available to the last issue which was available online on August 1, 2013. Some articles were not accompanied by abstracts, some abstracts were not in copyable formats, and many journals had not published online article abstracts for their early issues; however, the inclusion of a huge number of abstracts from 1967 to 2013 leaves little doubt concerning the representativeness of the corpus.

4.3 The software

For the purpose of investigating and listing frequent chunks of words in ALAAs, this study used AntConc 3.3.5, which was released in July 2012 (Anthony, 2012). AntConc is a freeware application which runs on both Windows and Linux systems. Although AntConc has a freeware license, it is easy to use and offers seven tools including a concordancer, word and keyword frequency generators, tools for n-gram analysis, and a concordance plot. AntConc Clusters/N-Grams tool displays word clusters and multi-word units or n-grams based on search conditions as shown in Figure 1.

Figure 1: AntConc Clusters/N-Grams tool



The N-Grams Tool searches a corpus for 'n'-length clusters and allows the user to find common expressions in the output. The size of n-grams can be set by the user and the software can display and order the n-grams alphabetically or by frequency. It is also possible to set a minimum frequency threshold for the displayed items.

5. Data Analysis

5.1 N-gram identification

As the software program AntConc 3.3.5 did not respond to a command for global cluster/n-gram analysis of the corpus due to overload, clusters ranging from two to nine words were obtained separately –which proved to be more convenient for study and further analysis. Moreover, because the inclusion of low frequency n-grams would have rendered the output lists unnecessarily long, the software was set not to produce n-grams with occurrences fewer than five. Thus, we had eight raw lists of word chunks or n-grams ranging from two to nine words with occurrences above five.

As a rule, the longer the chunks, the smaller the number of items in the output. In this corpus of 2,750,000 words, there were 60,617 types,

1,612,730 tokens of 2-grams, 43,723 types, 614,689 tokens of 3-grams, 14,507 types, 164,515 tokens of 4-grams, 3,851 types, 38,620 tokens of 5-grams, 1,079 types, 10,572 tokens of 6-grams, 417 types, 3,800 tokens of 7-grams, 162 types, 1,272 tokens of 8-grams, and 81 types and 591 tokens of 9-grams, all with frequencies higher than five. With this distribution of n-grams, it was inevitable that most of the chunks selected based on frequency should come from shorter chunks. Moreover, the more frequent the items, the smaller the number of the items with similar frequencies. For example, in the corpus, there was one 3-gram with 1,756 occurrences (*the use of*), one with 330 occurrences (*results suggests that*); there were two 3-grams with 325 occurrences (*results show that* and *there is a*), 15 with 90 occurrences, 48 with 50 occurrences, 340 with 20 occurrences, 1,983 3-grams with 10 occurrences, and 10,572 3-grams with 5 occurrences.

5.2 Short-listing the N-grams

The shorter n-grams of two to four words long in these lists were too numerous, and therefore impractical, to display in tables, even at a very high-frequency cut-off point — for example, there were 14,261 two-word n-grams (bigrams), 5,744 three-word n-grams, and 1,319 four-word n-grams occurring more than 20 times in the raw output. Moreover, the bi-grams were predominantly ordinary language expressions and most high-occurrence 3- and 4-grams seemed to be subsumable in longer ones. Only 5- to 9-word n-grams were selected so that space was saved and it was possible to include items with frequencies as low as 20, i.e., an occurrence rate of 2.27 per million words. This was a cut-off point much lower than the one used by Simpson-Vlash and Ellis (2011), who opted for 10 per million words as the cut-off frequency for the n-grams in their study, which according to them was among the lowest cut-off points in previous studies, used cutoff ranges between 10 and 40 instances per million words.

6. Results and Discussion

The goal of this study was to provide a profile of the common chunks and expressions found in a fairly comprehensive corpus of the abstracts of applied linguistics research articles. After examining and making decisions about cut-off frequency, the items in the lists of 5-,6-,7-,8-, and 9-grams were transferred to two master lists, one for content chunks, as a response to the first research question, and the other for meta-discursive or textual chunks, as a response to the second research question.

Table 1 displays content (non-textual) strings of five to nine words with frequencies above 20. Shorter clusters are not included because, at this cut-off point, they would make the list too long for an article. Besides, many of them are embedded in longer ones and, hence, are automatically reported, although the observed frequencies, which are usually much higher than their frequencies in the subsuming chunks, are not specified here.

In order to prepare Table 1, the items in the list of content n-grams were summarized by subsuming the shorter strings in the longer ones, removing the redundant ones. For example, the 5-gram *common European framework of reference* (36) was not included as it was part of the 6-gram *the common European framework of reference* (32). The reason for removing the shorter chunks in favor of the longer ones was the constraint of space. Although the exact frequencies of the subsumed chunks are not reported in the table, they can be safely assumed to be of much higher frequencies than the frequency reported for the longer ones.

However, this process of subsuming was not followed very strictly so that some overlapping chunks were left as examples to provide the readers with some information about the frequencies of chunks in subsuming relationship. For example, *English as a second language* occurs 478 times while *English as a second language (SLA)* occurs 58 times. Similarly , while *the test of English as a foreign language* occurs 34 times, removing article *the* yields a chunk with a frequency of 47 and adding the acronym *TOEFL* yields a 9-gram of 24 occurrences. *English as a foreign language* occurs 465 times, *English as a* 1132 times and *English language* 1,062 times. Hence, Table 1 can serve as a sample for extrapolation. This benefit can justify the many redundant cases which infiltrated the final list in spite of the criteria of subsumability.

There were very few 9-grams with high frequencies. In fact, only seven 9-grams had a frequency above 10. The most frequent ones had 24 occurrences. On the other hand, 5-grams were in the majority –in the absence of shorter n-grams, of course.

Table 1. Content n-grams longer than five words with frequencies above 20 in ALAAs

English as a second language478	English for academic purposes EAP82
English as a foreign language465	computer assisted language learning CALL76
in the foreign language classroom116	

non native speakers of english76
French as a second language63
in second language acquisition SLA58
learners of English as a second language54
native and non native speakers50
test of English as a foreign language47
second language acquisition SLA research46
the teaching of foreign languages42
learners of English as a foreign language38
in English for academic purposes38
in the second language classroom37
standards for foreign language learning35
the test of English as a foreign language34
learning English as a second language33
the common European framework of reference32
learning English as a foreign language32
Japanese as a foreign language31
Spanish as a foreign language29
the second language acquisition of29
in the teaching of english28
field of second language acquisition27
first and second language acquisition27
English as a lingua franca ELF27
the field of applied linguistics27
English to speakers of other languages26
the standards for foreign language learning26
for the teaching of english26
foreign language teaching and learning26
in computer assisted language learning26
the standards for foreign language26
to speakers of other languages26
a second or foreign language25
English for specific purposes ESP25
strategy inventory for language learning25
the zone of proximal development25
the test of English as a foreign language TOEFL24
on the teaching of foreign languages24
the field of second language acquisition24
people s republic of china24
Spanish as a second language24
in second and foreign language23
research in second language acquisition23
to examine the effects of23
approach to the teaching of22
German as a foreign language22
in the field of language22
English as a second language ESL learners22
in English as a foreign language EFL22
the ACTFL oral proficiency interview22
determine the extent to which21
in first and second language21
the development of second language21
the people’s republic of china21
students of English as a foreign language21
students of English as a second language21
by native speakers of english20
children with specific language impairment20
the study of second language20

Table 1 can provide pointers to some tendencies, give some insight, and raise some awareness. However, given the frequency-based selection

procedure and, hence, the absence of non-quantitative focuses in the n-grams, a comprehensive scrutiny of the items in the table is not attempted here and only some brief observations are made about selected features. As Table 1 shows, *English as a second language* (478) and *English as a foreign language* (465) top the list and are of similar frequencies. We may not know which authors have used the two chunks interchangeably and who have deliberately used *second* or *foreign* to draw a distinction; but, we can be reassured that both items are active and prevalent in the discourse of the field.

Some languages and nations occur more than 20 times in the corpus. This is due to their significant presence in research efforts and research reports. It should be noted, though, that shorter chunks are not included here, and we should not make sweeping generalizations based on this list. For example, *The United States*, *in the United States* and *in the USA* occur 343, 249 and 51 times, respectively. (N-grams including US were ignored because the software treated US and *us* indiscriminately.) Meanwhile, *in the United States the* occurs 30 times in the corpus but it was not included because it lacked an independent syntactic form.

It is acknowledged that chunks with lower frequencies than the cut-off of 20 may be equally or even more telling about the state of the art of applied linguistics because they are about more specific areas and concepts. Likewise, clusters such as *English as an additional language* (11) and *grounded reports of research and discussion of* (7) can usefully serve as fixed phrases in talking and writing about the respective topics.

Some concepts and ideas are represented by multiple words or expressions. For this reason, the frequencies of such items may be underestimated or they may not pass the specified threshold. For example, *in the field of applied linguistics, in applied linguistics field, in applied linguistics* are basically about the same thing. They may lose the race against an idea with similar or less prominence but more uniformity of expression.

Usually, there is a decline in the frequency of the n-grams as we move from shorter to longer chunks. But institutionalized phrases provide a challenge to this observation. Some longer n-grams may be as frequent, that is, some longer n-grams may be frequently used as highly lexicalized items. These long lexicalizations may justify occasional inclusion of such chunks as keywords in research reports. But, this is not to deny that the number of lexicalized short n-grams is far more. It may also be useful to note that there is an increase in the number of n-grams as we move down from more

frequent ones to less frequent ones in Table 1. While the highly frequent items at the top of the table are unique, there are more chunks with similar frequencies as we move down the table. This observation can be made about the other tables in this report or any other corpus–based list of lexical items in descending frequency order.

Table 2. Meta-discursive n-grams longer than five words with occurrences above 20 in ALAAs

the purpose of this study140	the results indicate that the38
this article reports on a117	implications of these findings are35
this paper reports on a103	the results indicated that the35
purpose of this study was102	this article focuses on the35
the purpose of this study was90	this paper reports on an35
the results of the study89	this paper reports the results35
the purpose of this study was to85	this study was to investigate35
the results showed that the70	this study investigated the effects of35
the results of this study66	this paper reports the results of34
the purpose of this article55	the present study investigated the34
the purpose of this paper55	the aim of this article33
the results of a study52	the present study investigates the33
the results show that the52	the results of a survey33
study investigated the effects of50	this article presents the results33
this article reports on the50	this article presents the results of32
this paper reports on a study50	this paper focuses on the31
the purpose of this paper is to49	the present study examined the28
the article concludes with a48	implications of these findings are discussed28
results are discussed in terms of48	reports the results of a study28
this article reports on a study48	this study was to examine27
purpose of this study is46	the present study was to26
concludes with a discussion of43	this paper presents the results of26
the purpose of this article is to43	this paper presents the results26
this article reports on an43	the aim of this paper25
the study reported in this article42	the implications of the findings25
the purpose of this study is to41	this article reports a study25
this article reports the results of40	the purpose of the present24
the implications of these findings40	these results are discussed in24
the findings of the study39	to investigate the effects of24
the findings of this study39	this paper reports the results of a24
the results suggest that the39	the implications of these findings for24
this paper reports on the39	the aim of this paper is to23
the aim of this study38	implications of the study are23
the purpose of the study38	

the findings of a study23	purpose of this study was to examine21
the findings suggest that the22	in this article we present20
the results are discussed in terms of22	the implications of the study20
were randomly assigned to one of22	the present study examines the20
this article reports the results of a22	this article is based on20
the findings are discussed in21	this article is concerned with20
these findings are discussed in21	this study investigates the effects20
this paper argues that the21	the first part of the paper20
the purpose of this study was to investigate21	this study investigates the effects of20

The textual or meta-discursive chunks occurring more than 20 times are listed in Table 2. As the size of the original list was not very large, all chunks which passed the frequency threshold of 20 were listed here, i.e., they were not summarized based on the criterion of inclusiveness. For example, *the purpose of this study* (140) is listed although it is included in *the purpose of this study was to* (85)/*is to* (43). Except for a few cases, e.g., *purpose of this study was to examine* (21), or *the present study investigates the* (33), the chunks in Table 2 are grammatically correct units. Longer n-grams are not in the table because they occur fewer than 20 times in the corpus. This is the reason the n-grams *the present study investigates the effects of* (10) and *the purpose of this study was to examine* (19), among other less frequent ones, are absent from this table.

As with Table 1, most of the clusters in Table 2 include five words because the cluster shorter than five words were not taken into account and the longer clusters predominantly fell short of the threshold of 20 occurrences in the corpus.

Among the most frequent verbs in the chunks reported in Table 2 are *reports* and the derivatives of be; and among the most frequent nouns are *article*, *paper,* and *purpose*. When checked in the list of bi-grams, *this article* ranked 10, *this study*, ranked 14, *this paper*, ranked 15, and *the article* ranked 73, while *the purpose* ranked 245. The words *article* and *paper* ranked 44 and 54, respectively in the frequency list of single words. This is an upshot of the obvious fact that shorter items embedded in chunks with certain frequencies are most likely to be of higher frequencies when considered separately because they may also be used in other combinations.

Moreover, there is widespread overlap and rather rampant crisscrossing: an n-gram with a certain frequency, e.g., *this study examines the relationship between* (12), embraces *this study examines the relationship*

(13), *study examines the relationship between* (20), *this study examines* (254), *this study examines the* (145), *the relationship between* (720), etc. Relatedly, in an n-gram such as *it is argued here that* (14), *here* is not an immediate part of the expression but is responsible for its low frequency because of the inability of the software to filter *here* out and add the frequency of the principal string to *it is argued that* (425+14). Following this line, one could take a further step and come up with proto-chunks or proto-n-grams based on a manual re-examination of the n-grams with or without the constraint of a lower threshold or cut-off point.

Table 2 shows that there are many ways to refer to the purpose of an article or its different elements. We should also note that there are many other less frequent meta-discursive points of departure to talk about an article or its content which do not feature in this table but are as effective and communicative as those reported, for example, *in this article we argue that* (9), *the article concludes by suggesting* (9), *these findings are discussed in light of* (7).

To compensate for the nonappearance of fairly frequent items in the list of meta-discursive chunks, which may practically be as useful as the frequent ones, and to add to the practical value of this research by further fine-tuning the portrait that it creates of the meta-discursive chunks in ALAAs, the meta-discursive expressions occurring more than 10 times were also identified and organized in bunches with a point of departure as a common element. Table 3 displays the results of this identification procedure. To prepare this table, the chunks which explicitly referred to the research report or its elements were selected and divided into two parts similar to the theme-rheme scheme in Halliday's (1985) functional grammar or the traditional topic-comment division, the first element serving as a given point of departure or springboard, and the second as new information, in a sentence or clause.

The n-grams longer than five words, which occurred more than 10 times in the corpus were examined and the items which included a verb and referred to the research report or its elements, e.g., *this article focuses on the, in this paper, I argue, the results of the study indicate,* were selected and listed alphabetically. In each selected n-gram there was a point of departure and a comment. The themes, which are usually shared by several rhemes, are bolded and arranged alphabetically in Table 3. The rhemes, which are usually verb phrases, were listed alphabetically under their respective themes or topics. Again, when there are competing rhemes/comments with

the same verbs and a frequency difference of less than five, e.g., *the article concludes with a discussion* (15), *this article concludes with a discussion of* (13), only the longer ones were kept to save space. This was part of the general strategy by this researcher, who was very strict at the n-gram extraction stage but, after becoming sure that few targeted items were left out, ready to compromise at the subjective stage.

It should be mentioned that the notion of theme/rheme is used only as a rough and ready organizing principle and is not followed in all theoretical details and technical aspects. For one thing, in many cases, the rhemes are curtailed because only the respective n-grams had frequencies higher than 10 and making the rheme part longer would mean chunks of lower frequencies.

Table 3. ALAA's points of departure (in bold) and their partial complements with occurrences higher than 10

in this article	was to16
i argue13	**the article**
i discuss10	concludes with a discussion of13
i examine13	concludes with a48
we discuss12	focuses on the12
we examine18	**the data**
we present20	were collected through12
we report18	**the findings**
in this paper	are discussed in terms of11
i argue10	are discussed in21
i examine the11	indicate that the18
i explore11	indicated that the11
we argue that15	showed that the10
we examine12	suggest that the22
present13	**the focus of this paper**
we report on13	is10
we report18	**the goal of this study**
we examined12	was to11
the aim of the present study	**the paper**
was to14	concludes with a discussion of the14
the aim of this article	concludes with a discussion20
is to31	concludes with a37
the aim of this paper	concludes with an10
is to23	ends with a13
the aim of this study	focuses on the10
is to17	**the present study**

aims to12
examined the28
examines the20
investigated the effects of10
investigated the34
investigates the33
is to16
was to investigate10
study was to26
the purpose of the present study
was to11
the purpose of the study
was to17
the purpose of this article
is to43
is50
the purpose of this paper
is to49
the purpose of this study
is to examine11
is to41
is45
was to examine the14
was to examine19
was to investigate the14
was to investigate21
was to85
was90
the results
are discussed in relation to11
are discussed in terms of22
are discussed in47
indicate that the38
indicated that the35
the results of the study
indicate that12
indicate14
show that16
the results of this study
indicate11
suggest that10
the results
revealed that the10
showed that both10
showed that the70
suggest that the39
the study
focuses on the11
found that the15
is based on15
shows that the13
was to determine10
was to investigate13
the study reported in this article
investigated10
these findings
are discussed in21
are discussed with12
have implications for10
suggest that the12
these results
are consistent with11
are discussed in terms of12
are discussed in terms12
are discussed in24
this article
argues that the13
concludes with a discussion of13
deals with the10
describes a study13
describes the development16
examines the role of10
focuses on the35
is based on a12
is based on20
is concerned with20
is to provide10
looks at the23
presents the findings of12
presents the results of a18
presents the results of32
provides an overview of11
reports a study of12
reports a study25
reports on a 117
reports on a study of14

reports on a study48
reports on an investigation13
reports on an43
reports on research10
reports on the results of10
reports on the50
reports the findings of a10
reports the findings of15
reports the results of a22
reports the results of an11
reports the results of40
this paper
argues that the21
focuses on the31
is based on12
is concerned with17
looks at the14
presents the results of a17
presents the results of26
presents the results26
reports on a study of11
reports on a study that11
reports on a study which11
reports on a study50
reports on a103
reports on an35
reports on the39
reports the findings of15
reports the results of a study17
reports the results of a24
reports the results of34
sets out to10
this study
examined the effects of16
examines the effects of10
examines the relationship between12
focuses on the16
investigated the effect of17
investigated the effects of35
investigated the effects36
investigated the relationship between11
investigates the acquisition10
investigates the effect of11
investigates the effects of20
is to examine17
was designed to18
was to determine17
was to examine the19
was to examine27
was to explore13
was to investigate the22
was to investigate35

6. Conclusion

This research creates empirically derived lists of formulaic sequences of words frequently used in the abstracts of articles reporting applied linguistics research. It offers lists of frequent n-grams used for textual and meta-discursive purposes as well as content n-grams which well coincide with recurrent institutionalized expressions representing major current ideas and conceptual balance in the field. So, the output, modest though, can be of descriptive and pedagogical value and promote awareness about the phraseology of a very important and thriving academic text type.

The study of prefabricated, established “formulaic chunks of language, resulting from memorizing the sequences of frequent collocations” (N. Ellis, 2003, p. 68) seems all the more worthwhile when we relate it to the theory of chunking in psychology, as formulated by Miller (1956). Miller proposed the concept of chunking as “an extremely powerful weapon for increasing

the amount of information that we can deal with" (p.93). This is in harmony with Ferguson's (1994) idea that there is a tendency for human language to become conventionalized at various levels. That is, people in recurring contexts and situations tend to use a limited range of utterances as memory friendly "templates" (Gobet, 2005). Lewis (1997, 2000), who is closely associated with "lexical approach", has also put much emphasis on lexical chunks and collocations. In both of these works he emphasizes raising language learners' consciousness about lexical chunks and conventional phrases in order to extend their proficiency.

Several researchers have empirically shown the benefits for learners of raising awareness about formulaic language (e.g., Boers & Lindstromberg, 2006; Gatbonton and Segalowitz, 2005). In a recent article Boers and Lindstromberg (2012) reviewed experimental and intervention studies on formulaic sequences published since 2004 and concluded that learners gain a lot from building a sizable repertoire of L2 formulaic sequences. Their suggestion to language teaching practitioners was to draw learners' attention to formulaic sequences, encourage them to use corpus tools, and helping them commit particular formulaic sequences to memory.

The applied linguistics research reporter, like any other writer, should allow for both creativity and for the established patterns and text schemata, which, according to Hyland (2007) often form the basis of any variations. As hinted above, cognitive science has shown that formulaic expressions crucially pave the way for fluent processing. This research has contributed to the understanding of some of the textual conventions of ALAAs by making available some of the most basic prefabricated phrases. It has relevance for the teaching of research writing to NNS and it can help the novice researchers to write discipline-specific reports and their teachers to prepare basic instructional materials to teach the novice the word sequences which are essential to producing ALAAs a la genre.

This paper did not engage in a detailed and fine-tuned discussion of the listed items as the goals were broad-based and no particular grammatical or conceptual features were targeted and the identified items were therefore inclusive of diverse functions and concepts. Discussion of specific functions and features require more focus and the in-depth analysis of particular items or categories of items, which in turn, requires a different research design. This study focused on the frequent lexical chunks in ALAAs. It will be instructive to study the occurrence of these chunks in other academic disciplines or to juxtapose them with lists based on corpora from those

disciplines. Furthermore, the analysis of the frequent points of departure and their complements in ALAAs is rough-tuned. A more fine-tuned profile of the points of departure for talking about the articles and their contents needs further analysis including shorter n-grams and pre- and post- modifiers of the main objects of writing, i.e., *article, paper, study, research, etc*. and the secondary level words, i.e., results, findings, implications, etc. Exploring shorter and less frequent lexical chunks will certainly provide additional collocational awareness about ALAAs, or other text types for that matter. One can also take a sociocultural approach and analyze the meta-discoursive chunks within the functional framework that Kuhi and Behnam (2011) created to identify the interpersonal and contextual forces behind textual choices.

References

Anthony, L. (2012). AntConc Program (Version 3.3.5) [Computer Software]. Retrieved June 6th, 2013 from dshttp://www.antlab.sci.waseda.ac.jp/

Biber, D., & F. Barbieri. (2006). Lexical bundles in university spoken and written registers. *English for Specific Purposes*, *26*, 263-286.

Bloch, J. (2010). A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research, 2*(2), 219-244.

Boers, F., & Lindstromberg, S. (2012). Experimental and Intervention Studies on Formulaic Sequences in a Second Language. *Annual Review of Applied Linguistics, 32*, 83-110.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238.

Croft, W., & A. Cruise. (2004). *Cognitive linguistics*. Cambridge: Cambridge University Press.

Davis, A. (2007). *An introduction to applied linguistics*. Edinburgh: Edinburgh University Press.

Del Saz, M.M. (2011). *A pragmatic approach to the macro-structure* and metadiscoursal features of research article introductions in the field of Agricultural. *Sciences English for Specific Purposes*, *30*(4), 245-318.

Ellis, N. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C.J. Doughty and M.H.

Long (Eds.). *The handbook of second language acquisition* (pp. 63-103). Oxford: Blackwell.

Farjami, H. (2013). A corpus-based study of the lexical make-up of applied linguistics article abstracts. *The Journal of Teaching Language Skills*, 5(2), 27-50.

Ferguson, C.A. (1994). Dialect, register and genre: Working assumptions about conventionalization. In D. Biber and E. Finegan (Eds.). *Sociolinguistic perspective on register* (pp.15-30). New York: Oxford University Press.

Flowerdew, J. & M. Peacock, (Eds.). 2001. *Research perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review, 61*(3), 325-53.

Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology, 19*, 183-204.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Halliday, M.A.K. (1966). Lexis as linguistic level. *Journal of Linguistics, 2*(1), 57-67.

Halliday, M.A.K. (1985). *An introduction to functional grammar*. London: Edward Arnold.

Hartley, J. (2003) Improving the Clarity of Journal Abstracts in Psychology: The Case for Structure. *Science Communication 24*(3), 366–79.

Hawes, T., & Thomas, S. (1997). Tense choices in citations. *Research in the Teaching of English, 31*(3), 393-414.

Hinkel, E. (2004).Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research, 8*(1), 5-29.

Huckin, T. (2001). Abstracting from abstracts. In M. Hewings (Ed.). *Academic Writing in Context* (pp. 93-103). Birmingham: University of Birmingham Press.

Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*(2), 156-177.

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor: University of Michigan Press.

Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes, 27*, 4-21.

Jalilifar, A. R. (2010). The status of theme in applied linguistics articles. *The Asian ESP Journal*, *6*(2), 7-39.

James, C. (2007). What is applied linguistics? *International Journal of Applied Linguistics*, *3*(1), 17-32.

Kuhi, D., & Behnam, B. (2011). Generic variations and metadiscourse use in the writing of applied linguists: A comparative study and preliminary framework. *Written Communication, 28*(1), 97-141.

Lewis, M. (1997). *Implementing the lexical approach: Theory into practice*. London: Language Teaching Publications.

Lewis, M. (2000). *Teaching Collocations*. London: Language Teaching Publications.

Lim, J. M. H. (2006). Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes*, *25*(3), 282-309.

Lorés, R. (2004). On RA abstracts: From rhetorical structure to thematic organization. *English for Specific Purposes*, *23*(3), 280-302.

Malaskova, M. (2012). Hedges as writer protective devices in applied linguistics and literary criticism research articles. *Discourse and Interaction, 5*(1), 31-47.

Martin, P. M. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes*, *22*(1), 25-43.

Martinez, R., & Schmitt, N. (2012). A phrasal expression list. *Applied Linguistics*, *33*(3), 299-320.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.

Nation, I.S.P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press.

Nattinger, J. R., & J. DeCarrico. 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press.

Pho, P. Z. (2008). Research article abstracts in applied linguistics and educational technology: A study of linguistic realizations of rhetorical structure and authorial stance. *Discourse Studies*, *10*(2), 231-250.

Samraj, B. (2005). An exploration of genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes*, *24*, 141-156.

Santos, M. B. (1996). The textual organization of research paper abstracts in applied linguistics. *Text, 16*(4), 481-499.

Shin, D., & P. Nation. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal 62*(4), 339–48.

Simpson-Vlach, R., & N. C. Ellis. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics 31*, 487–512.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. London: Routledge.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.

Tseng, F. (2011). Analysis of move structure and verb tense of research article abstracts in applied linguistics. *International Journal of English Linguistics, 1*(2), 27-39.

Ventola, E. (1994). Abstracts as an object of linguistic study, in S. Cmejrkova, F. Danes & E. Havlova (Eds.). *Writing vs. speaking: Language, text, discourse, communication*. Proceedings of the conference held at the Czech Language Institute of the Academy of Sciences of the Czech Republic, Prague, 14-16 October, 1992 (pp. 333-52). Tubingen: G. Narr.

Yang, R., & Allison, D. (2003). Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes, 22*(4), 365-385.

Appendix

Language Teaching Acquisition
Language Teaching Research
Language Testing
Learning and Instruction
RELC Journal
Research in the Teaching of English
Second Language Research
Studies in Second Language
System
TESOL Quarterly
The Annual Review of Applied Linguistics
The ESP Journal
The international Journal of Applied Linguistics
The Modern Language Journal
Applied Linguistics
Canadian Modern Language Review
Classroom Discourse
Computer Assisted Language Learning
ELT Journal
English for Specific Purposes
English in Education
English Teaching_ Practice and Critique
Foreign Language Annals
International Journal of Multilingualism
Journal of English for Academic Purposes
Journal of Second Language Writing
Language Acquisition
Language Awareness
Language Learning
Language Learning & Technology