Iranian Journal of Applied Linguistics (IJAL), Vol. 16, No. 2, September 2013, 1-19

Author-assigned Keywords in Research Articles: Where Do They Come from?

Esmat Babaii [a]
Assistant Professor of Applied Linguistics, Kharazmi University, Tehran, Iran

Yoones Taase
MA graduate of TEFL, Kharazmi University, Tehran, Iran

Received 5 June 2013; revised 4 July 2013; accepted 25 August 2013

Abstract

The current study attempts to explore the characteristics of author-assigned keywords in research articles as important constituents of targeted search in academic communities. To this end, the keywords of 200 research papers in the field of applied linguistics, in terms of domain, degree of specificity, and relation to the titles, were analyzed. To supplement the findings, the keyword choice strategy of a number of researchers with publishing experience in the field was also investigated. The analysis revealed a considerable rate of title-keywords match, especially with respect to field-specific keywords. This finding points to the importance of users' field-specific background knowledge in locating relevant information on the web. The examination of authors' viewpoints and strategies, on the other hand, helped to bring to light the complex and non-clichéd nature of keyword selection. The significance of authors' diverging and converging attitudes and their implications for enhancing the success rate of keyword search are discussed.

[a] *Email address*: babai@khu.ac.ir
Corresponding address: Department of Foreign Languages, Kharazmi University, No. 43, Mofatteh St., zip code 15719-14911, Tehran, Iran

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

Keywords: Indexing; Keyword selection; Author-assigned keywords; Scientific internet search; Research article title; Bibliographic metadata

Introduction

Being a successful student or an academic in the digital age would require information-seeking skills that could match the massive amount of scientific information provided by Web technology. One of these skills is internet search to gain fine-grained access to the target information; a process which is not always as felicitous as expected. In fact, research into the nature of on-line information seeking strategies suggests that users are not generally good at following the models used in existing search engines (see Muramatsu & Pratt, 2001; Teevan et al., 2004). According to Teevan et al. (2004), people tend to do the search through two major strategies: 'orienteering' and 'teleporting'. They define orienteering as "situated navigation, [where the users implement] a series of small steps to narrow in on the target, without specifying the entire information need up front" (pp. 415, 417). Teleporting, in contrast, is defined as an attempt on the part of the users "to jump directly to their information target using keywords" (p. 417). The second strategy, in fact, is encouraged and supported by search engines which operate on the basis of 'indexing' the documents, a process applied to "the content of documents to select those concepts that best represent them, and thus facilitate storing and retrieval" (Gil-Leiva & Alonso-Arroyo, 2007, p. 1175). Indexing involves selecting keywords or key-phrases. Howcroft (2007) defines keyword as follows:

> A keyword (also known as index term or descriptor) is a term that captures the essence of the topic of a document or a search query. It is used to retrieve documents in an information system […]. It can consist of a word, phrase, or alphanumerical term. Keywords are created by analysing the document either manually with subject indexing or automatically with full text indexing or more sophisticated methods of keyword extraction (p. 75).

As stated by Ercan & Cicekli (2007), the main difference between automatic and manual procedure—which is done by the author(s) of the documents or library

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

cataloger(s) — is that in the former "most indicative phrases in a document are selected as keyphrases for that document. Thus, automatic keyphrase extraction algorithms are limited with phrases that appear in the text" (p. 1706), whereas manual keyphrases might or might not occur in the text. In fact, 'keyness' of keywords in automatic indexing depends on their statistically unusual high frequency in the document (cf. Dunning, 1993; Scott, 2001, 2005) but a human agent— an author or a library cataloguer— may have other criteria in mind.

While the field of information technology is replete with studies on automatic keyword extraction techniques (for instance, Turney, 2000; Shah et al., 2003; Wu et al., 2006; Ercan & Cicekli, 2007; Nguyen & Kan, 2007; Wu & Li, 2008; to name but a few), little is known about the nature of author-assigned keywords as a type of manual indexing. As a matter of fact, manual indexing is somehow dismissed because of being time consuming and costly (Wu et al., 2006). Nevertheless, author-assigned keyword selection, if done systematically, has this potential to increase the success of keyword search (Mauer et al., 2011) as the keywords supplied by authors have an important presence in database descriptors (Gil-Leiva & Alonso-Arroyo, 2007).

Among notable research on author-assigned keywords, reference can be made to Kipp (2005), Gil-Leiva & Alonso-Arroyo (2007), Heckner et al. (2008), Strader (2009), and Schwing et al. (2012). Comparing the indexing practices of authors, users, and trained indexers on a social bookmarking site, i.e., Citeulike, Kipp (2005) finds notable differences between the three groups in terms of both frequency and type of the terms employed for tagging academic articles. Detailing the differences, she advocates the inclusion of user tagging in the design of the information systems: "user tagging, with its lower apparent cost of production, could provide additional access points to traditional controlled vocabularies…" (p. 435).

In another study, Gil-Leiva & Alonso-Arroyo (2007) analyze several database records to examine the presence of author-assigned keywords. They find a rate of 46% presence of these keywords either as exactly the same or in the nominalized form. They conclude that "keywords provided by authors are a valuable source of information for both human indexing and for automatic indexing systems of journal articles" (p. 1181). Heckner et al. (2008), studying the tagging behavior of users and authors' keyword assignment, attempt to explore the functional and

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

linguistic characteristics of keywords. They find a low rate of match (30%) between users' tags and authors' keywords. They conclude that taggers avoid specific terminologies and employ simple, general words, which is in contrast to author's choice of field-specific terms which they postulate to be a "differentiation strategy with respect to a possibly huge amount of literature in the same field" (p. 13).

Strader (2009) attempts to examine the degree of overlap as well as non-overlap (or uniqueness) between author-assigned keywords and cataloguer-assigned Library of Congress Subject Headings (LCSH) for a set of electronic theses and dissertations in Ohio State University's online catalogue. Matching the keywords and LCSH against the dissertation titles and abstracts, she finds a complementary relationship between the two methods of assignment. In a partial replication of Strader's (2009) study, Schwing et al. (2012) report a high rate of match between title, abstract and keywords. They justify the observed match by stating that "the abstract, keywords, and title are all provided by the author, so they are bound to share the same vocabulary and structure" (p. 919). The degree of non-match, accordingly, is due to the fact "the [author-assigned] keywords tend to represent more current, cutting edge ideas, as well as terms that are more specific within the sciences, LCSH, in contrast, tends to be more stable and to connect to broader subjects (p. 924, emphasis added). The interesting point about Schwing et al.'s (2012) study is the way they interpret the findings. They do not consider uniqueness or lack of term-overlap as a liability but an asset: unique terms have this potential to provide more access point to the documents.

The above-mentioned investigations have been able to initiate a new line of research in information technology where different perspectives of the involved parties and their possible effects on the outcome of online communication are at stake. To make a contribution to this under-researched but flourishing area of scientific information management, the present study attempts to explore: (a) the nature of author-assigned keywords in research articles in terms of domain-specificity and relation to the titles, and (b) the researchers-authors' viewpoints and strategies in keyword selection. In particular, it intends to provide an enriched understanding of keyword selection by incorporating an actor's perspective to a corpus analytic methodology. This line of research, hopefully, can enhance the quality of communication between the suppliers and users of scientific information.

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

Methodology

The current study involved two phases of corpus analysis and collecting information through an online open-ended questionnaire. The details of both phases are provided below.

Corpus Selection

A corpus of 200 research article abstracts were chosen randomly from four celebrated English medium journals in the field of applied linguistics including *System*, *Language and Communication*, *Journal of English for Academic Purposes* and *English for Specific Purposes*. It was intended to select journals published by the same publisher (here Elsevier Science Publication) so that possible effect(s) of different editorial policies are controlled. Inclusion of author-assigned keywords was the most important selection criterion and the date of publication was limited to 2000-2011.

Online Questionnaire

In order to explore the viewpoints and keyword selection strategies of academics with publishing experience, an open-ended questionnaire was designed. The questionnaire intended to add a new dimension to the investigation through including 'actor's point of view' (Davis, 1995) in the study of author-assigned keywords. In fact, the questionnaire helped to cater for an "interpretive orientation, in which the immediate (often intuitive) meanings of actions to the actors involved are of central interest" (Erickson, 1986, p. 120). This ethnographic approach can be justified by considering the fact that the academics with publishing experience have been (and continue to be) involved with the process of keyword selection. Therefore, the nature of keywords is directly influenced by their perceptions and attitudes. It would be particularly informative to find areas where their viewpoints diverge and converge. The researchers-authors participating in the study were requested to provide detailed answers to the followings questions:

To what extent keywords are important to you, why?
When you are going to publish your articles in journals, how do you select your keywords?
What is the effect of your article's title in selecting the keywords?
How do you decide about the number of keywords?
What is the effect of terminology of your field in selecting the keywords?

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

What is your preference in keyword selection: general terms or field-specific ones? Why?

Procedure

For the corpus analysis phase of the study, the words in the titles and the keyword sets were analyzed. It should be noted that in counting the words in the titles, the word counting method introduced by Buxton & Meadows (1978) and recommended by Bachir & Buxton (1991) was used. In this method, the words that carry no useful information—called 'stop words'—such a prepositions, conjunctions, articles, and pronouns are not included in the counting. To determine the generality or field-specificity of the words, we consulted the most recent edition of *Longman Dictionary of Language Teaching and Applied linguistics* (Jack C. Richards & Richard Schmidt, 2002) and *Encyclopaedic Dictionary of Applied Linguistics* (Keith Johnson & Helen Johnson, 1998). In the following example, regular type words are considered as specialized terms and the underlined word is considered as general:

Teaching styles, learning styles, <u>success</u>, Grasha-Riechmann learning style scale, Teaching Styles inventory

To ensure the accuracy of the procedure, the analysis was done independently by the researchers and a high degree of inter-coder agreement was obtained (phi=.89). For the second phase of the study, to encourage a high rate of cooperation, the questionnaire was sent to about 100 academics we knew in person. However, only 40 academics provided detailed responses to our questions. Incomplete cases were discarded. Fortunately, despite relatively small size of the sample, there existed variations among the respondents in terms of the country of origin, location of their institutions and years of experience. The detailed analysis of the responses is presented in the next section.

Results and Discussion

In this section, first the results of the analysis of keywords and titles will be detailed. Then, the participants' responses will be presented and discussed.

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

Titles and Keywords, Domain and Match

The results of the quantitative analysis of keywords in terms of domain and their relationship to titles are presented in Table 1 below. As it is observed, the total number of content words in the titles was 1399 with an average of 6.9 per article. The observed frequency of author-assigned keywords was 955 with an average of 4.8 per article.

Table 1

The frequency and type of words in the keywords and titles

	Total	**Min**	**Max**	**Mean**
Keywords	955	2	10	4.8
General keywords	203	0	9	1.1
Specialized keywords	752	1	8	3.8
Words in title	1399	3	20	6.9
Keywords in title	414	0	5	2.1
General keywords in title	80	0	4	0.4
Specialized keywords in title	334	0	4	1.7

Min=minimum number, Max= maximum number

Out of 955 keywords, 414 cases were repeated in the titles, creating a 43% match or overlap between titles and keywords. However, a careful study of the observed frequencies reveals that keyword-title match is mostly due to the presence of specialized terminologies rather than general terms as the ratio of specialized keywords in title to the total number of keywords in title is 334/414= 80.7%. The high incidence of specific terms in keywords points to the fact users need to be equipped with domain-specific background knowledge to find the needed information in the articles. In other words, if users employ only the general terms in their queries, it is unlikely to lead to a successful search as the ratio of general terms, whether in titles or in keywords is low (80/1399=5.7%, and 203/955=21%, respectively).

This conclusion is supported by the findings of Hölscher & Strube (2000) who compared the successful search behaviour of experts with the relatively unsuccessful search of newbies. Their study indicates that with the help of domain-specific knowledge, experts find the needed information with less time spent and significantly fewer of terms in their queries (average query length: 1.9 vs. 2.9

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

words). To put it another way, the privileged 'insider' status of experts is also revealed in their internet search through the use of jargon which is by definition "speech or writing used by a group of people who belong to a particular trade, profession or any other group bound together by mutual interest […] A jargon has its own set of words and expressions which may be incomprehensible to an outsider" (Richards & Schmidt, 2002, p. 278). This implies that attempts to optimize electronic access to the needed information in scientific publication should be multidirectional. That is, as we are trying to improve the qualities of databases through advancements in web technology, we should not neglect the efforts to educate the users regarding the strategies they can employ to increase the efficiency of information search and management.

Keywords as Perceived by Authors

As stated before, the second phase of the study attempted to uncover researchers-authors' perceptions about the nature of keywords. In what follows, the responses are presented and discussed.

In response to the first question—*To what extent keywords are important to you, why?*—all the respondents, except two who happened to be well-known scholars in the field, stated that keywords were quite important. As for the reason, however, different explanations were offered.

Most of the respondents found the keywords important for relatively general, commonsense reasons, mainly irrelevant to the role of keywords in information technology, while about 30% of them revealed their awareness about internet search techniques and mentioned their concern about findability of their article: a concern about "how widely it is read, reviewed and cited by others to enhance the 'impact' of their research" (McDonnell, 2010, p. 5). Responses (1) and (2) represent the first mentality (a commonsense view towards the importance of keywords), whereas (3) and (4) are more reader-oriented by highlighting the importance of retrievability.

(1) They are important, as thinking about them obliges me to focus on what's really important.

(2) They are quite important because, along with the title and abstract, they are representative of the content and theme of my paper.

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

(3) They are helpful tools to locate very specific information usually in search queries […] as they can be considered 'research address'…

(4) Of course, I want to maximize the 'presence' or impact of my research – keywords are a useful way to 'tip the scales' in my favour.

There was one interesting response (No. 5) which demonstrated a deeper knowledge about the inner relationships in the academia: things are different for a novice researcher and a well-known academic. Response No. 5 could also help us understand why the importance of keywords was less strongly pronounced by experienced researchers in our sample. In fact, because of being highly cited in the field, renowned scholars did not probably feel much need for more visibility in electronic databases which could be enhanced through appropriate use of keywords for their published works.

(5) They are important depending on where you are publishing. If you are a well known author publishing in a top journal, your articles will probably be located in other ways. However, if you are not well known and publishing in a lower journal that is sufficiently indexed, your keywords are very important. Next would be your title followed by your abstract.

As for the second question—*When you are going to publish your articles in journals, how do you select your keywords?*—to most of the respondents, the relevance of keywords to the theme of the paper was a crucial selection criterion. Responses No. 6 and 7 neatly formulate the idea. Some (50%) indicated that the theme was naturally reflected in the article title, hence, making a relation between title and the keywords inevitable (see No. 8 and 9). Assuming a natural link between the theme and the title is in line with Bachir & Buxton (1991) who consider this a common expectation from the authors to indicate the subject matter in the title. However, they also remind that some authors provide "attractive but unindicative title (p. 59) to win the readers' attention", a practice they regard as improper.

About one third of the respondents expressed their concern about maximizing the number of successful hits for search engines (see No. 10), adhering to the

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

notion of 'keyness as frequency' used in search technology (cf. Dunning, 1993; Scott, 2001). A small number (10%) of respondents mentioned that they would search for a model in the works of other researchers (No.11). Finally, 5% of participants revealed that they did not use any particular strategy to do the job (see No. 12)

(6) I think, it's more a question of "relevance"; since a paper can be subsumed under a certain area/subfield, the general but relevant word describing the paper comes first (e.g., reading comprehension) followed by the more specific and relevant words (e.g., assessing reading FOLLOWED by reading assessment techniques)

(7) I try to select the content words which display the underlying theme of my paper…

(8) I put down the words that come to my mind first; often they are also included in the paper's title.

(9) I select my keywords basically according to the content of my article, but they are practically embedded in the title and abstract of the article.

(10) I think about my main themes and write a keyword which I judge to be the most frequently used in my context, in an attempt to maximize the number of successful hits.

(11) I usually look at the keywords used by established authors who are publishing similar research. I also look at the keywords used in recent theoretical books/articles related to my research in order to get a better idea of what is current.

(12) I'm not systematic. I simply brainstorm on …

The third question explicitly inquires about the effect of title on the choice of keywords. Most respondents (about 80%) acknowledged a relationship between the

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

title and the keywords, but not a causal one. Response (13) lucidly explains the link between theme/content, title and keywords:

(13) I refer to the whole paper when writing the title and selecting the keywords. I mean I do not select the keywords on the basis of the title. However, since the title like keywords should represent the content of the paper, naturally some of the key words are mentioned in the title as well.

In fact, as mentioned by the respondents a 'partial match' is to be expected but some authors did care about 'uniqueness' of the keywords (to use Schwing et al.'s 2012 terms) as well (see No. 14)

(14) I don't want to repeat too many keywords in the title, but there is usually one or two that are important to have in the title and in the keyword list.

About 20% of the respondents, nevertheless, denied a relation between the title and the keywords. To them, the functions of these two are quite different: titles are to attract the readers and the keywords are to enhance the article's findability, (see No. 15 and 16):

(15) No effect really; the goal of the title is to attract the reader's interest, while the search words are to maximize successful hits from search engines.

(16) I don't think the title should reveal ALL that is in the paper. There's something aesthetic about choosing the title. However, keywords are meant to facilitate internet search.

Ironically, this is in sharp contrast with the common librarian advice which promotes informative value of the title in scientific publication and expects the authors to restrain their desire for attracting attentions through catchy titles (see the quotation below).

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

> The editor of scientific or scholarly material looks for a title that *defines* and *delineates*. A journal editor, like a book editor, looks for informative value in the title. Words in titles *should be indicative of what the publication is about*. *Catchy titles that draw attention but do not index well should be reserved for popular magazines and trade books which will be less used by serious researchers*. If an author sees merit in such a title, it can counterbalanced by a subtitle with more substantive keywords. Titles should either provide a succinct rendering of what the publication is about or provide a statement suggesting inquiry or conclusion. A good rule is that the most significant keywords in the abstract should be in the title. (Davis, 1997, p.20, emphasis added)

However, this craving for attention-grabbing titles, as also documented by Haggan (2004) in the fields of linguistics and literature, seems to be much more common than imagined. Such observations could provide a good opportunity to study areas of conflict between personal tastes and disciplinary norms.

The responses to the fourth question—*How do you decide about the number of keywords*?—were also varied, despite our expectations. While 60% of the respondents considered no choice for the authors in this respect, (see No. 17), others mentioned their own strategies (No. 18, 19). Interestingly, the magic number seemed to be 5 (see No. 20, 21, and 22).

(17) As for me, the easiest question is the one about the number of keywords because that is usually pre-determined by the publisher (through the electronic submission system) and it's very unusual for authors to get a choice.

(18) I usually try to write as many as possible, again to try to maximize hits.

(19) I have no specific rule. I usually select between 4 to 10 keywords, maybe because I've seen that papers published in well-known academic journals have the same number of keywords. Moreover, 4 to 10 keywords have always been enough to represent the main theme of the papers I've written so far.

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

(20) I generally choose five unless there is some compelling reason to include one or two more.

(21) Normally, the keywords should be no more than five.

(22) […]. Even in case there is not a limit by a given editorial board, I do not exceed 5 since this would mean violation of some unstated rules!

In response to the fifth question—*What is the effect of terminology of your field in selecting the keywords?*—the respondents unanimously emphasized the important role of terminology (see No. 23, 24, and 25). Only one respondent doubted its significance (No. 26).

(23) The keywords often tend to be the terminology of the related field of study, maybe because they are usually the representative content words. Moreover, I think there is an inclination to select specialized terminologies as keywords so as to sound academic.

(24) This is crucial! Without understanding the language of the field, there is little hope that anyone in the field will be able to find your article with a keyword search.

(25) Terminology can be useful for searches because it narrows down the search field on the internet and, hopefully, allows interested readers to find my articles more easily.

(26) It might have an effect; it depends on the nature of the research, I guess. If it's more exploratory, there might not yet be established terminology in the field.

Considering the significance of field-specific terminologies for authors, one can conclude that successful queries should include such terms as their indispensable ingredients. The corpus analysis phase of this study, along with Hölscher and Strube (2000), provides further support for this conclusion.

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

The last question attempted to explore the nature of keywords in terms of knowledge domain: the respondents were asked to say whether they gave priority to general or field-specific keywords. Here, the consensus was on a 'balance' or a 'mix' of both. In 50% of the responses, there was no indication of prioritizing generality or specificity (see No. 27). Within this group, only 5% indicated adopting a systematic general-to-specific strategy for their keyword selection (see No. 28 and 29).

(27) I think the generality of the keywords is as important as their specificity. I believe that keywords should help someone who has not read the paper decide whether the paper is the one he is looking for or not. Therefore, the keywords should be general enough to enable him distinguish the related field of study and specific enough to let him know whether what he is looking for exactly is covered in the paper or not.

(28) Keywords are research addresses, so I arrange them from general to specific. Yes, this order is quite important to me. I've read nowhere that keywords should be arranged like this but I've done so in my few published papers. Interestingly, I've received no comments from reviewers for my arrangements.

(29) My selection is like an 'inverted triangle' that narrows down: Discipline/Field, Field/Method, and Area of Research/Research question/Genre. This way I give 'a big picture' of my research by situating it in a reasonably broad disciplinary framework that may help non-experts look up the jargon terms coming next, and at the same time I economize editorial space and address fellow researchers specialized in the field.

Among those who mentioned their priority, 40% treated 'specificity' as more important, since it helped to make their research stand out among other similar studies, what Heckner et al. (2008) call 'differentiation strategy' (see No. 30 & 31). However, a small minority (10%) said they would prefer to use general terms to attract a wider audience (No. 32).

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

(30) Specificity, because otherwise there is little chance to show how the article differs from any others in the field.

(31) It would be most desirable to have a list of keywords that demonstrates key characteristics of the article. Therefore, a healthy balance of both is more desirable. That being said, if one approaches the notion of keywords as a distinguishing factor, it would be most desirable to lean towards the specificity along the continuum.

(32) I think having more general than specialized keywords is beneficial. You want to maximize the potential your article can be located by a wide variety of readers. However, you need a few specialized keywords so that those looking for more specific topics can find your research.

Concluding Remarks

The research reported in this paper, despite its admittedly limited scope, helped to illuminate some less commonly acknowledged aspects of keyword assignments in academic publishing. The corpus analysis phase of the study verified the significant presence of specialized terminologies of the field in both the keywords and the title, suggesting the usefulness of educating novices to enhance information search success through expanding their domain-specific knowledge.

The second phase of the study, on the other hand, confirmed the value of an ethnographic approach in the field of information processing and management where success of communication depends on understanding the motives and hidden agenda different actors bring to the situation. To highlight the important findings of this phase, reference can be made to the following points:

(I) While the significance of keywords was generally acknowledged by the participants, well-known experts seemed to treat keyword selection less seriously than expected. The observed variation in the responses, of course, could not be simply reduced to expert-novice differences but, as reported in previous research (e.g., Hölscher & Strube, 2000; Salager-Meyer, 2001; and Koutsantoni, 2006), the

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

degree of expertise may influence a researcher's attitude, strategy use and discursive practice;

(II) In selecting the keywords, authors use quite diverse strategies: some use the paper's theme as a guide, some consider a potential audience, and some check relevant articles by well-known authors; some follow a systematic approach while some jot down whatever comes to their mind. In short, keyword assignment is far from being a simple, standardized process;

(III) Authors show disagreement as regards to the relation between the title and the keywords. To most of them, there is a natural link between the two while to a visible minority title selection is considered an opportunity to catch the reader's attention, what Bhatia (2007) calls communicating 'private intention' within the context of organizational culture which assumes that the sole purpose of the title is to inform about the content of its corresponding academic article (see Bachir & Buxton, 1991 and Davis, 1997);

(IV) Most authors prefer a balance between generality and specificity in their selection of keywords to attract a broad audience but not at the expense of specialist readers. In fact, the balance is tipped in favour of domain-specific terminology as this practice serves the double purpose of improving search efficiency as well as revealing the author' technical expertise within a particular research area. These findings point to the intricate and discursively dynamic processes involved in this seemingly plain routine in academic life.

Research into the nature of author-assigned keywords may be furthered through investigating (a) possible effects of disciplinary variation on the keyword selection process, especially the contrast between hard and soft sciences, and (b) the moderating effect of cultural norms in non-English contexts.

Acknowledgements

The corresponding author wishes to thank the research deputy at Kharazmi University, Tehran, Iran for funding the sabbatical leave at Macquarie University, Sydney, Australia during which some parts of the study was done. Thanks are also due to all researchers-authors in the field who took part in the study and shared their valuable ideas and strategies with us.

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

Notes on Contributors

Esmat Babaii *is a lecturer of applied linguistics at Kharazmi University, Iran, where she teaches research methods, language assessment and discourse analysis to graduate students. She has served on the editorial boards and/or the review panels of several international journals. She has published articles and book chapters dealing with issues in Systemic Functional Linguistics, test-taking processes, and critical approaches to the study of culture and language.*

Yoones Taase *Holds B.A of English Language and Literature from University of Semnan and M.A of TEFL from Kharazmi University. He has been an EFL teacher of Kish Institute of science and technology since 2010. Multiple intelligences is his main research interest and he has published an article in PAAL journal 2012.*

References

Bachir, I., & Buxton, A. (1991). The information content of titles of Arabic periodical articles. *Journal of Information Science, 17*, 57-63.

Bhatia, V. (2007). *Worlds of written discourse: A genre-based view*. London & New York: Continuum.

Buxton, A., & Meadows, A. (1978). Categorization of information in experimental papers and their authors abstracts. *Journal of Research Communication Studies, 1*, 161-182.

Davis, K. (1995). Qualitative theory and methods in applied linguistics research. *TESOL Quarterly*, *29* (3), 427-453.

Davis, M. (1997). Title Keyword Selection and Use for Optimum Document Retrieval. *Public & Access Services Quarterly*, *2*(2), 15-22.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19*, 61-74.

Ercan, G., & Cicekli, I. (2007). Using lexical chains for keyword extraction. *Information Processing and Management*, *43*, 1705-1714.

Erickson, F. (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 119–161). New York: Collier-Macmillan.

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

Gil-Leiva, I., & Alonso-Arroyo, A. (2007). Keywords Given by Authors of Scientific Articles in Database Descriptors. *Journal of the American Society for Information Science and Technology, 58*(8), 1175-1187.

Haggan, M. (2004). Research paper titles in literature, linguistics and science: dimensions of attraction. *Journal of Pragmatics, 36*, 293-317.

Heckner, M., Mühlbacher, S., & Wolff, C. (2008). Tagging tagging: Analysing user keywords in scientific bibliography management systems. *Journal of Digital Information*, *9* (2).

Hölscher, C. & Strube, G. (2000). Web search behavior of Internet experts and newbies. *Computer Networks, 33*(1-6), 337-346.

Howcroft, G. (2007). A Beginner's Guide to Metadata and Keywords. *Editors' Bulletin*, *3* (3), 75-77 .

Johnson, K., & Johnson, H. (1998). *Encyclopedic dictionary of applied linguistics*. Blackwell Publishers Ltd.

Kipp, M. E. I. (2005). Complementary or discrete contexts in online indexing: A comparison of user, creator, and intermediary keywords. *Canadian Journal of Information and Library Science, 29*(4), 419-436.

Koutsantoni, D. (2006). Rhetorical strategies in engineering research articles and research theses: Advanced academic literacy and relations of power. *Journal of English for Academic Purposes, 5*, 19-36.

MacDonnell, K. (2010). The key to keywords. *The Journal for Specialists in Group Work*, *35*(1), 3-6.

Maurer, M., McCutcheon, S., & Schwing, T. (2011). Who's Doing What? Findability and Author-Supplied ETD Metadata in the Library Catalog. *Cataloging & Classification Quarterly*, *49* (4), 277-310.

Muramatsu, J., & Pratt, W. (2001). Transparent queries: Investigating users' mental models of search engines. *Proceedings of SIGIR '01*, 217-224.

Nguyen, T., & Kan, M. (2007). Keyphrase Extraction in Scientific Publications. *ICADL*, 317-326.

Richards, J. C., & Schmidt, R. (2002). *Longman dictionary of applied linguistics and language teaching* (Third Edition). Pearson Education Limited.

Salager-Meyer, F. (2001). From Self-highlightedness to Self-effacement: A genre-based study of the socio-pragmatic function of criticism in medical discourse. *LSP & Professional Communication*, *1*(2), 63-84.

Schwing, T., McCutcheon, S., & Maurer, M. (2012). Uniqueness matters: Author-supplied keywords and LCSH in the library catalog. *Cataloging & Classification Quarterly*, *50* (8), 903-928

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]

Scott, M. (2001). Comparing corpora and identifying key words, collocations and frequency distributions through the WordSmith Tools suite of computer programs. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. 47-67). Amsterdam: John Benjamins.

Scott, M. (2005). Wordsmith Tools 4.0. Oxford: Oxford University Press (Available at: www.lexically.net/wordsmith/ version4). (accessed February 2013).

Shah P., Perez-Iratxeta, C., Bork, P., & Andrade, M. (2003). Information extraction from full text scientific articles: Where are the keywords? *BMC Bioinformatics*, *4*, 20-33.

Strader, R. (2009) Author-Assigned Keywords versus Library of Congress Subject Headings: Implications for the Cataloging of Electronic Theses and Dissertations. *Library Resources & Technical Services, 53* (4), 243-250.

Teevan, J., Alvarado, C., Ackerman, M., & Karger, D. (2004) The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. *CHI Letters*, *6* (1), 415-422.

Turney, P. (2000). Learning Algorithms for Keyphrase Extraction. *Information Retrieval, 2*, 303-336. Kluwer Academic Publishers.

Wu, Y. B., & Li, Q. (2008). Document keyphrases as subject metadata: incorporating document key concepts in search results. *Information Retrieval, 11*, 229-249.

Wu, Y. B., Li, Q., Bot, R., & Chen, X. (2006). Finding nuggets in documents: A machine learning approach. *Journal of the American society for information and technology (JASIST)*, *57*(6), 740-752.

[Downloaded from ndea10.khu.ac.ir on 2026-11-07]