J. TECHNICAL WRITING AND COMMUNICATION, Vol. 41(2) 139-160, 2011

OUTLINING PURPOSES, STATING THE NATURE OF THE PRESENT RESEARCH, AND LISTING RESEARCH QUESTIONS OR HYPOTHESES IN ACADEMIC PAPERS

WASIMA SHEHZAD
Yanbu University College, Kingdom of Saudi Arabia

ABSTRACT

Driving research questions from the prevailing issues and interests and developing from them new theories, formulas, algorithms, methods, and designs, and linking them to the interests of the larger audience is a vital component of scientific research papers. The present article discusses outlining purposes or stating the nature of the present research, and listing research questions or hypotheses in the introduction of academic papers. This corpus-based genre study focuses particularly on Move 3 of the model "occupying the niche." The results indicating disciplinary variation show that the writers of Computer Science (CS) research articles, over the years have developed an increased use of outlining purpose/stating the nature of the present research, having the characteristics of purposive, descriptive, extension of the previous work, contrast to the existing work, brevity, complexity, and a description of methodology. It also shows that listing research questions or hypothesis may have distinctively different functions in developing genres as compared to the established ones such as physics.

INTRODUCTION

Listing research questions in academic papers leads the readers in the specific direction, as, to use Heidegger's (1954) words in Lovitt (1977, pp. 3-35), "Questioning builds a way," at the end of which researchers' investigations and findings

doi: 10.2190/TW.41.2.c
http://baywood.com

attain prominence, significance, and relevancy to the burning issues of their discourse community. Scholars are expected to drive research questions from the prevailing interests related to the real world matters and to the epistemic world of research, and develop from them new theories, formulas, algorithms, methods, and designs, linking to the interests of the larger audience. Thus, the present study investigates the significant issues of outlining purposes and addressing research questions/hypotheses in writing the introduction of a research article. The article focuses on Computer Science (CS) which is one of the most popular subjects of today. Unfortunately, scholarly studies in this discipline such as Cooper (1985), Posteguillo (1995, 1999), Kuo (1999), and Anthony (1999) have been mostly based on the structure of the Introduction. Second, these works have either been too general or relied on a small corpus, creating a need for a deeper analysis based on a larger corpus. Hence, this exploratory study investigates various strategies and styles employed by Computer scientists in explaining the nature of their own research and raising the research questions in terms of Swales and Feak's (1994, 2004) CARS (Create A Research Space) model.

Swales (1990) and Swales and Feak's (1994, 2004) CARS model which is based on three rhetorical moves provides a structural framework for writing the introduction of a research article. Move 1 establishes research territory by showing that the research area being handled is important, interesting, and worth researching, and reviews previous works in the area (see Shehzad, 2006). Move 2 establishes the niche by indicating shortcomings or "gaps" in the already published research (see Shehzad, 2008). "The role of Move 3 is to turn the niche established in Move 2, into the research space that justifies the present article" (Swales, 1990, p. 159). This is done by filling the created gap, answering the specific question, or continuing the rhetorically established tradition. In the final move of a typical research paper Introduction, an offer to fill the gap (or extend the tradition) that has been created in Move 2, is made. Lewin, Fine, and Young (2001) also consider it as the most salient interpersonal component of Move 3 and opine that authors do this either explicitly or implicitly "by carefully echoing or paraphrasing the gap." Move 3 is typically signaled by some reference to the present text (this, the present, reported, here). In addition, Lindeberg (1994) has used the term "boost" for the explicit promotion of the research contribution element in Introductions.

According to Swales and Feak (2004) the major purpose of Move 3 (which is obligatory) is to outline purpose or state the nature of the present research. Along with the investigation of this step, the present article discusses the second step of Move 3 "listing research questions or hypotheses" which, according to the CARS model, is present in some fields and is optional. Contrary to this, the present findings suggest obligatory nature of this step in CS. Both of these steps co-occur physically and rhetorically in the Computer Science Research Articles' (CS RA) Introduction.

LITERATURE REVIEW

CS, one of the most intellectual developments of the 20th century, changed the entire conception and outlook to the world around us in the 21st century. The research article is one of the major genres to provide access to the globalization of scientific findings and occupies an important strategic position in scientific writing. It is ironical that the number of research articles published in CS is remarkably increasing every year but linguistic investigations in this area are quite limited.

In today's busy world, the introduction serves the same purpose as a headline in a newspaper. The more interest, strength, validity, and relevance it holds, the more audiences it is going to attract. Introduction is a means of "putting him [reader] in the picture"; therefore, "reasons for undertaking the investigation or for writing the article" should be explained (Godfrey & Parr, 1960, p. 53). Although Booth (1993, p. 4) does not emphasize stating the question, but neither undermines its importance, "the introduction should state the problem, and perhaps ask a question. The objective must be clear." Before setting the scene in detail, one must answer the question: Why did you do the work and what is its purpose? Along with stating the purpose of the article, rationale for the study should be summarized (O'Connor, 1991). Similarly, Barrass (1978) advocates that the reason why the work was undertaken, indicating the scope of the work, should be explained in the Introduction. Blicq (1986, p. 237) also emphasizes the inclusion of "a definition of the problem and the specific purpose of the investigation." Introduction can help the writers to "feel grounded with a good sense of direction" as it "sets the scene for the reader and indicates the direction of the rest of the piece" (Crème & Lea, 2003, pp. 113, 114). However, no one provides a scaffolding as integrated and complete as Swales (1990) and Swales and Feak's (1994, 2004) CARS model for building around the Introduction of a research article. It has been divided by them into three moves, and Move 3 being the focus of this study will be addressed here.

The first step of the Move 3 was called "a kind of promissory statement" by Swales (1990, p. 159), as the writer makes a "contract" with the reader to deliver something "and the formulation of this may reflect the degree of author commitment in various ways," Lindeberg (2004, p. 41). In Swales' (1990) CARS model it is presented in two steps:

- The author or authors indicate their main purpose or purposes.
- The author or authors describe what they consider to be the main features of their research.

These two steps have been combined into one step in the CARS model (Swales & Feak, 1994, 2004). This is called Move 3 Step a. The first element in Move 3 (i.e., *outlining purpose or stating the nature of the present research*) is obligatory (Lewin et al., 2001; Swales & Feak 2004). The former divides this

into two variants: Purposive, the author or authors indicate their main purpose(s); and Descriptive, the author or authors describe the main feature of their research. About the second step "listing research questions or hypotheses," Swales and Feak (1994, 2004) believe, it is probable in some fields. Since "each professional community has its own set of genres, with varying degrees of uniqueness" (Orr, 1999, p. 32), Move 3 Steps a and b, as found in the CS RAs Introductions, have been analyzed in this article, not only to determine their probable, possible, or obligatory nature, but also to understand what strategies are used in the realization of these steps.

METHODOLOGY

"A corpus is a large collection of authentic texts that have been gathered in electronic form according to a specific set of criteria," as Bowker and Pearson (2002, p. 20) advise. Based on the criteria proposed by Reppen and Simpson (2002), a corpus (Shehzad Computer Science Corpus) was created which was representative as it included almost equal number of articles from the five major journals in the target field and also because of the high academic standing of the publisher. The corpus also covered a wide range of topics. Native and non-native authors' articles were not differentiated based on Okamura's (2003, p. 51) claim that, "working in the same discipline such as science, writers share the norms and expectations about the use of linguistic forms, regardless of their nationality and language."

Texts for the present study were based on the Introductions of 56 research articles published in five different journals of the prestigious Institute of Electrical and Electronics Engineers (IEEE). These included: IEEE Transactions on Computer Science (ToC); IEEE Transactions on Pattern and Machine Intelligence (PAMI); IEEE Transactions on Software Engineering (SE); IEEE Transactions on Parallel and Distributed Systems (PADS); and IEEE Transactions on Knowledge and Data Engineering (KDE). Details of the creation of the corpus have been accounted in Shehzad (2007a, 2008, 2010). The corpus was studied quantitatively through WordSmith tools (Scott, 1997) and Swales and Feak's (1994, 2004) CARS model was used for rhetorical analysis.

RESULTS AND DISCUSSION

Discussion of the results has been divided into two sections. After a comparison with other similar studies, I discuss salient features of Move 3 Step a, as found in CS, including its characteristics of being purposive, descriptive, contrastive, brief, complex, and methodology oriented. This is followed by a discussion of the deictic references used in the employment of these elements. The first part concludes with a comment on the tense and purpose statement.

1. Move 3 Step a: Outlining Purpose or Stating the Nature of the Present Research

A comparative analysis of this step with other similar studies in the field of CS was done. The current figure (see Table 1) falls between the figures of the two earlier studies of 1995 and 1999, depicting the near mandatory nature of this step.

I fully agree with Swales' (1990) assertion about the writers' obligation for writing something about the purpose or nature of the present research. An introduction without this step seems like having nice toppings of the "Announcement of Principal Findings" and "Value statements" etc., but lacking the basic ingredient. Out of the 56 Introductions in the SCS Corpus, one did try to venture without this (see Appendix). In this introduction, a two-page literature review is all inclusive of Move 1, Step a, Move 1, Step b, and Move 2. The rest of the two brief paragraphs are about the contribution their work makes to the research sphere, followed by an outline of the structure of the article.

Salient Features of Move 3 Step a

Here, I will discuss the salient features of Move 3 Step a, including but not limiting it to its "purposive" (author(s) main purpose(s)), and "descriptive" (description of the main features of research) nature (Swales & Feak, 2004, p. 262). The elements of extension, contrast, brevity, complexity, and methodology specific to CS rhetoric have also been explored.

Purposive—The purposive nature (an indication of the main purpose or purposes of the research) of this step was also reported by Posteguillo (1999) who specified 25% occurrence of this step in his corpus. The results of the present study demonstrate even lower use of the purposive statements. Nevertheless, with

Table 1. Comparative Studies on Move 3 Step a

Studies	CARS	Percentages
Posteguillo (1995)	3.1.a 3.1.b (Swales, 1990)	25% 95%
Anthony (1999)	3.1.a 3.1.b (Swales, 1990)	41.7% 100%
Present study (2009)	Move 3 Step a (Swales & Feak, 1994, 2004)	98.21%

12.50%, it provides evidence of its availability as an option for Computer scientists, though less frequently exercised. Example:

> In this paper, we propose Hamming norm computation as a basic operation of interest for data stream processing. . . . *There are two compelling reasons* for proposing Hamming norms. *First*, in the. . . . *Second*, when we apply Hamming norm to. . . . (INTR 53)

One of the Introductions even has a heading that says, "Goals and Outline of the Paper." Such exceptions indicate that the need to write about the purpose of research does exist somewhere and may increase with the passage of time.

For outlining purpose, the writers also use the lexical items such as *goal, objective,* etc. The SCS Corpus had the following hits as shown in Table 2.

Some examples of these indicators are given below.

- GOAL
  - Our *goal* is to provide a set of characteristics that are . . .
  - Our *goal* is to detect and represent . . .
  - . . . this *goal* is elusive due to . . .
  - Toward this *goal*, we propose . . .
  - . . . to achieve this *goal*.

So when Computer scientists mention their goals, they are very explicit in informing the reader what they want to achieve.

- PURPOSE

  "In research texts, 'studies' and not authors have a purpose," Lewin et al. (2001, p. 52) inform. This was also noticed in the present work, though it is not true for aim(s) and objective(s) where both authors and studies may be subjects. There is no example of authors having a purpose here, but some other things definitely have, such as:

Table 2. Lexical Indicators of Outlining Purpose in CS RAs Introductions

Lexical items	Parts of speech	Occurrences
Goal	Noun	16
Purpose	Noun	3
Aim	Verb	3
Aim	Noun	2
Aims	Verb	1
Objective	Noun	12
Objectives	Noun	3
Total		40

– The primary *purpose of an object recognition system is* to dynamically interpret imagery data to determine . . .
– *The purpose of this study* is descriptive in a certain context . . .
– . . . *The purpose of a scenario specification* is to describe how a system is intended to behave.

The definiteness of having a purpose and clearly stating it is obvious in their lexical selection (i.e., through heavy use of the words such as purpose, aims, and objectives).

- AIM/AIMS
  Having a goal enables the researcher to focus and devise ways to achieve it.
  – In this paper, we *aim* to provide . . .
  – In this way, we *aim* to support . . .
  – . . . with the *aim* of developing . . .
  – . . . with the *aim* of estimating . . .
  – This work *aims* at creating a . . .
- OBJECTIVE/OBJECTIVES
  These examples of having clear goal posts reflect high level of objectivity by setting the objectives that are measurable.
  – . . . is important in terms of our more general *objective*, which is to . . .
  – The main *objective* of a . . .
  – Our *objective* is to develop an efficient . . .
  – Further, since our *objective* is to maximize the . . .
  – . . . control policies for redirection are other key *objectives* of this work.
  – This implies a host of consequent *objectives*, listed below:
  Sometimes some secondary statements, such as "In addition . . . ," "Additionally . . . ," "A secondary aim . . . ," and "A further reason for . . . ," are also introduced to complete Move 3 Step a.

Having discussed the features of the Purposive element of Move 3 Step a, I'll now move to its Descriptive element.

Descriptive—". . . the writers are responsive to rhetorical pressures, rarely commencing with a simple purpose statement, but typically displaying an attempt to situate a claim in an important area of scientific endeavor" (Hyland, 1998, p. 367). Most of the CS Introductions' writers prefer to write about the nature of the present research in descriptive form. Some of the descriptions of the features of the research are fairly long. These are the ones that also include information about the method(s) and technique(s) used. The authors declare that they have carried out an investigation in order to fill the knowledge gap created in the earlier move. In this declaration, the methods, justification for the methods, the hypotheses and the results may be included. One of such Introductions offered a sub-heading for this part that said "1.2. Overview."

Extension of previous work—Computer science is not always about the invention of new models. The existing models are also sifted and refined by using either new or alternative techniques. Consequently, in the SCS Corpus, 7% of the authors' work is about the extension of the previous work. Examples:

> *In extension to the previous* unified architectures, *our design incorporates* a third type of operation with the MM algorithm: the polynomial multiplication in NTRU's Public Key Cryptosystem, which is simply the usual convolution product of two vectors. (INTR 4)

> *In this paper, our work is primarily confined to an enhancement to the original protoco*l and no change in the original protocol is required. (INTR 7)

> *This paper extends our previous work* [2], [3], [1] by combining a number of methods into a unified framework that is applicable to real world problems. (INTR 14)

> *This paper is an extension of our paper* [21]. *We have added extensive* experimentation on a large number of sequences taken from commercial movies and the Web. (INTR 22)

This enhancement and enlargement is not only made in the contributions of their fellow scholars of their community but many a times their own previous findings are reinforced through its extension in their present research as is seen in the last two examples, above.

Contrast to existing work—In the backdrop of the nature of the present research, the need to document the contrasting factor between and among two/various research works is observed in the CS discourse. The contrast in CS Introductions is drawn on the bases of various factors, such as:

- specific assumptions;
- existing approaches;
- extension in the concept;
- database languages;
- models;
- applications; and
- implementations.

In one Introduction, almost one page is devoted to differentiate the present work from the related work previously done. Example:

> *Our study differs from the previous research efforts* mentioned above *in a number of aspects. First of all, our study introduces* a unique approach for handling . . ., and an architecture for. . . . The database and knowledge-based systems are. . . . *Second,* as the language of the architecture. . . . *Third, we treat* imprecise data and knowledge inherent . . . by utilizing a . . . model. *In addition, we model* uncertainty and. . . . *Fourth, we introduce* an extension to virtual class concept with fuzzy rules. Users can. . . . *Moreover*, a virtual class

> can inherit from. . . . The reason to add this capability to the environment is to satisfy . . . requirements . . . and to facilitate . . . more efficiently. *Finally, in this study, we also focus on* the implementation issues . . . architecture. . . . (INTR 58)

In this example, transitions of sequence, additional support, and conclusion signal the logical relation between a number of contrasting aspects of the present and the previous disclosures.

Brevity—Shehzad (2007a) reports that the CS RAs Introductions are fairly lengthy; however, when it comes to the realization of Move 3 Step a, it is found that 16% of the Introductions have very brief Move 3 Step a. Example:

> This paper presents an improved, systematic method to convert march tests for BOMs into march tests for WOMs; it is based on [4]. Subsets of this problem have been addressed in [5], [6]. (INTR 10)

> In this paper, performing a simulation study, we investigate whether MMRE is a reliable selection criterion or not. The findings suggest … (INTR 35).

These instances of short and crisp moves support CARS model and can serve as good examples for pedagogical purposes. Nevertheless, the factor of complexity as described below needs to be accorded as well.

Complexity—Despite the above exemplification of brevity, most of the Move 3s in CS Introductions are complex and lengthy. Sometimes it is hard to differentiate between Move 3 Step a and Step c which is describing the nature of the present research and reporting the major findings. Example:

> This motivated us to develop a novel approach that integrates detection and tracking into a unified framework—the temporal approach. It uses the temporal relationships between. . . . The proposed algorithm is divided into two phases—Phase I (detection) and Phase II (prediction and update-tracking). Phase I-Detection. This phase detects the regions of interest that potentially contain faces. . . . This phase also identifies the detection parameters. . . . These parameters can be computed using frame-by-frame detection. However, the detector response can decrease due. . . . Without any additional information, these responses can easily be rejected even if they indicate the presence of a face. It is therefore important to incorporate the temporal information in a video sequence, which we do in Phase II with a prediction-update model. Phase II-Prediction and update. In this phase, we use a Condensation filter and factored sampling to propagate the detection parameters over time and, hence, track the faces through the video sequence. The local maxima of the probability distribution produced by the detectors, indicate possible occurrences of faces and are used to initialize the procedure. Samples are picked with a Gaussian centered at each maximum. These samples with their corresponding detection probabilities are propagated over time using a zero order prediction model on the detection parameters. The prediction of detection parameters guarantees the accuracy of accumulation

> of detection probabilities and continuous detection. This helps to stabilize detection and to make it less sensitive to thresholds compared to the frame-based detection procedure. The tracker integrates the response from the two detectors with the temporal information to detect intermediate poses. We also handle the appearance of faces by updating the tracker with the probabilities produced by the detection routine. (INTR 22)

Such occurrences are not only complex but can sometimes be confusing. The reason for this complexity could be the peculiar nature of some research work or lack of inclination to follow a particular writing style. The manuscript guidelines provided by IEEE journals do not include any advice on writing the introduction.

Outlining purpose or describing methodology?—Discussion of methodology, a significant structural stage of any scientific work, is one of the few important points missing from CARS model (Swales, 1990; Swales & Feak, 1994, 2004). Some authors of CS RA prefer to write the details of the method(s) used in the research rather than stating the purpose or describing the nature of their research. Probably, the focus of their research is the novel method(s) adopted, developed or designed. Examples:

> *In this study, we present a method* for recovering an illumination distribution of a scene from image brightness observed on an object surface in that scene. . . . (INTR 15)

> In this work, we investigate in a descriptive empirical study the relationship of the inspection factors team size, reading techniques, and inspection duration with the. . . . *We employ a statistical model* from recent research [3], which uses the empirically measured probability distribution of defects found by a population of inspectors, who used a *particular inspection technique*, to compute the defect detection effectiveness, cost, and benefit of hypothetical nominal teams. A hypothetical nominal team consists of an arbitrary combination of average inspectors from a set of inspector populations, e.g., inspectors using one from a set of defect detection techniques. *For the empirical study, we collected time-stamped defect data* from inspectors who read up to eight hours in a controlled experiment. We investigate the effort used for a given reading technique, the effect of actual inspection duration on the effectiveness of a given reading technique for single inspectors and teams, and determine the optimal mix of defect detection techniques for a nominal team of given team size and given optimization criteria using a statistical model. (INTR 29)

In the first example, devising and presenting the method itself is a big achievement. Similarly, detailed descriptions of the methodology explained in the second example, having phrases such as "we employ a statistical mode . . .," "used a particular inspection technique," "collected time stamped defect data," and "determine the optimal mix of defect detection techniques," suggest a need for the inclusion of another step, namely description of methodology, in Move 3 of CARS for CS.

Deictic References

The onset of Move 3 Step a is often marked by the use of deictic references to the present text such as: *this, the present, we, reported, here, now, I, herein,* etc. (Swales & Feak, 2004). The following deictic references were noted in the Introductions of CS research articles. These deictic references along with the number of hits in SCS Corpus and their referents are given in Table 3.

Current seems to be the more popular choice of the writers than *the present* in the use of Move 3 Step a. In the CS RA Introductions there were ten entries of *current* for this rhetorical stage. However, in the remainder of the articles, the deictic *current* referred to the nouns given in Table 4 in general:

Authors also commonly switch at this stage from impersonal to personal by using the pronoun *we,* etc. The heavy use of the personal pronoun *we*, to give prominence to the authors' personalities in CS RA Introductions, has been discussed by Shehzad (2007b). According to Swales and Feak (2004, p. 263), "these signals come early in the sentence." For example, of the 48 introductions in the Swales' (1981) corpus, only in one article, the deictic *In this paper* phrase occurred at sentence-final position. On the contrary, out of the total hits for the phrase *In this paper* in the Introductions of SCS Corpus, 60% came in the beginning of the sentences as suggested by Swales and Feak (2004) whereas out of the remaining 40%, 12% were near beginning, 13% occurred at the end of the sentences, and 15% in the middle of the sentences. Some examples of

Table 3. Deictic References in CS RA Introductions

Deictic	Hits	Comments
Current	20	3 referred to paper, research, and work; and the rest to the technical matter.
Study (noun)	16	Present research
Reported	10	4 referred to the present work and 6 to others work
Now	8	
Here	4	
Research	3	This research, current research, the focus of our research
The present	2	Referred to paper and work
Report	2	Present research

Table 4. References of the Deictic *Current*

Deictic	Reference
Current	approaches (2)
Current	approximations
Current	association rule miners
Current	commercial active database system
Current	cryptosystems
Current	density
Current	dependencies
Current	language environments
Current	measurement
Current	models
Current	multicast tree
Current	paper
Current	research
Current	state
Current	static network model
Current	systems
Current	targets
Current	work
Current	and a next state in the system

the uses and different placement of the phrase *In this paper*, from the Corpus are given here:

- At the Beginning:
  - *In this paper*, we focus on a new hardware-based prefetcher.
  - *In this paper*, we derive a tight lower bound on the guaranteed throughput of a multizone disk that. . . .
  - *In this paper*, we introduce a novel optimization technique, which is based on. . . .
- Near Beginning:
  - The emphasis *in this paper* is on the other component of tracking. . . .
  - The encoding proposed *in this paper* also gives us a formal framework for. . . .

- At the End:
  - A production quality real-time video processing application for obstacle detection is used as a case study *in this paper*.
  - It also serves as a discussion justifying the work presented *in this paper*.
- In the Middle:
  - . . . more closely related to the problem described *in this paper* (though their focus is not really on. . . .
  - . . . is an important and well-researched problem, to improve response times and system utilization, the work that is described *in this paper* is quite different since it deals with mapping (scheduling) the tasks within a single job (in this case. . . .

In the examples where the deictic, *In this paper* has been used in the beginning, the writers' personality is foregrounded contrary to its occurrence at the end of the sentences where it is the paper, the study, the work that gets importance. Whereas, the occurrences in the middle of the sentences suggest the prominence being given to the method, technique, or the particular research point.

We can infer from these examples that whenever the phrase *in this paper* is used in the beginning of the sentence it is complimented by we + verb and we get phrases like those given in Table 5.

Tense and Purpose Statement

Swales and Feak (2004, p. 263) advise using the present tense when referring to the *type of text* (paper, article, thesis, report, research note, etc.) and to use either *was* or *is* for referring to the *type of investigation* (experiment, investigation, study, survey, etc.). The analysis of the concordance lists retrieved from SCS Corpus showed the results given in Tables 6 and 7.

Considering the results shown in Tables 6 and 7, Swales and Feak's (2004) suggested relationship between tense and purpose statement seems to be fluid. They propose the use of present tense for the *type of text,* but we see about 10% use of the past tense with it in the CS RA Introductions. Similarly, the option of the use of *was or is* for the *type of investigation* is pretty open. Their idea needs further investigation before it can be theorized or used pedagogically, at least in Computer Science.

Move 3 Step b Listing Research Questions or Hypotheses

In this section, after the identification and comparison of disciplinary and interdisciplinary variation, salient features of this step have been discussed.

"The main purpose of the Introduction is to provide the rationale for the paper, moving from general discussion of the topic to the particular question or hypothesis being investigated," Swales and Feak (2004, p. 222) claim, yet hedge it

Table 5. Examples of We + Verb Structure with the Phrase *In this paper*

Phrase	We + verb	No. of entries
In this paper	We propose	5
In this paper	We present	3
In this paper	We address	3
In this paper	We focus	2
In this paper	We introduce	2
In this paper	We describe	2
In this paper	We show	2
In this paper	We refer	1
In this paper	We develop	1
In this paper	We consider	1
In this paper	We explore	1
In this paper	We investigate	1
In this paper	We use	1
In this paper	We argue	1
In this paper	We follow	1
In this paper	We derive	1
In this paper	We revisit	1

by saying that this step is probable in some fields and not in others. Of the four Introductions that they considered, only one had this element. Lewin et al. (2001), on the other hand, propose that Move 3 often lacks a statement of hypotheses or research questions. In this regard, the results of the present study are compared in Table 8 with the similar works done earlier.

Quite contrary to Posteguillo's (1999) findings, in Anthony's (1999) corpus, this step that he called *Question raising*, was missing. However, the figure in the present study is slightly higher than the figure reported by Posteguillo (1999) but still less than Swales and Najjar's (1987) figure about Physicists—32.14% occurrence of listing research questions or hypothesis in the CS RA Introductions reflects an occasional use of this step by the Computer scientists indicating its low tendency in developing sciences. Inter-disciplinary variation among the journals is reported in Table 9.

PAMI and SE seem to include some Introductions having research questions and hypotheses in them. These are the two journals that have Introductions with clear research questions. On the other hand, the figure of KDE is negligibly low.

Table 6. Tenses Used with the *type of text*

Type of text	Simple present	Simple past	Present perfect	Comment
Paper	37	1	1	
Article	—	3	—	
Report	—	—	—	
Research				Used as verb
Total	37	4	1	

Table 7. Tenses Used with *type of investigation*

Type of investigation	Simple present	Simple past	Present perfect
Study	10	1	—
Experiment	—	2	—
Investigation	1	1	—
Total	11	4	—

Table 8. Comparative Studies about *listing research questions or hypotheses* Step

Studies	Subject	Percentage
Swales and Najjar (1987)	Physics Education Psychology	45% 7%
Posteguillo (1999)	Computer Science	22.50% (Question raising)
Anthony (1999)	Computer Science	zero (Question raising)
Present study	Computer Science	32.14%

Table 9. Journal-wise Occurrence of Move 3 Step b

Journals	Percentages
IEE Transactions on Computer	5.35%
IEEE Transactions on PAMI	8.92%
IEEE Transactions on SE	8.92%
IEEE Transactions on PADS	7.14%
IEEE Transactions on KDE	1.78%

Salient Features of Move 3 Step b

The significant features of Move 3 Step b based on the Corpus can be listed as: research questions, assumptions, and hypothesis.

Research questions—Out of the 32% occurrences of listing research questions or hypothesis, 7% Introductions of CS RAs offer clear research questions in the Corpus which is quite a low figure to put any claim about its usage. Nevertheless, the ones having questions look like the following examples:

> Consequently, *the important question* concerning the approximation of the problem arises: *How good is* a computed minimizer relative to the unknown global optimum? *Can a certain* quality of solutions in terms of its sub-optimality be guaranteed in each application? (INTR 23)

> *A basic question is whether* defect detection during inspection is an individual activity, where inspectors find most of the defects before or instead of a meeting, or rather a group activity, where most defects are detected during discussion in a meeting. (INTR 29)

> However, complete specifications are typically not available at the start of a design analysis phase (Move 2). Instead, we begin with some key properties and attempt to use the model checker to validate them. *When the properties do not hold, and they seldom do, what is at fault, the properties or the design?* (INTR 34)

The only Introduction that lists questions under a sub-heading is from Software Engineering. It enumerates eight questions along with their explanations, spread over one page. Example:

> *Research Questions*
>
> *Does distributed* work introduce delay, as compared to same-site work? . . .
>
> *What factors* influence the time interval required to make a software change?

> *What role*, if any, does spreading work across multiple sites play in lengthening this interval? Assuming that there is an association between distributed work and longer intervals, there are many distinct ways in which working across multiple sites might introduce this delay. . . .
>
> *What differences are there* between same-site and distributed social "networks" and their effectiveness? . . .
>
> For example, *what is the relative size of* local and distributed social networks?
>
> *Is there a perception of* greater misunderstanding of tasks, priorities, plans, and changes across sites?
>
> *To what degree is work* at the different sites interdependent?
>
> *Does interdependence* diminish over time? (INTR30)

Whether this example is just an exception or could be used as a role model needs further investigation.

Assumptions—Unlike hypothesis (that may be the choice of many scientists) Computer scientists prefer to make assumptions. There are 7% cases, some examples of which are given here.

> Obviously, *we can assume* without loss of generality that each batch consists of exactly n requests as we are interested in the worst-case scenario. (INTR 11)
>
> Rather than schedule multiple jobs, *we are assuming* that the NOW is dedicated for our vision application and we are trying to assign the tasks of this application to the different NOW nodes, which is itself a very hard problem. (INTR 38)
>
> *We assume* that k _ n, and that the data are sparse i.e., the average number of 1s per row is substantially smaller than the number of attributes. (INTR 59)
>
> *The assumptions that we made* for the proposed approach are as follows: Known geometry. . . . No interreflection. . . . (INTR 15)

Computer scientists are as much interested in writing about their assumptions as they are in writing about their research questions.

Hypothesis—Although there are 25 concordance hits for the word *hypothesis* and 16 for the *hypotheses*, they are not used in the sense linguists would normally use them, for example; "A set of mutually exclusive and exhaustive hypothesis alternatives is called a hypothesis," has been used in one Introduction. There is only one good example found in the Corpus which is given in the following example:

> *Our hypothesis* was that by obtaining a better estimate of the optimal cover required for determining the box dimension, the fractal dimension estimates would become more accurate. (INTR 13)

In the rest of the 13% cases of making hypothesis, the writers tend to put up their thesis in an implied manner. Four such examples are given in the following examples:

> *Given a* fixed host density, *we can* easily determine a best threshold to use. (INTR 5)

> Our focus in this paper is to explore how such prior knowledge might be combined with more general grouping expertise, within a modern probabilistic framework. (Move 3 Step a) We argue that there are many practical scenarios in both human and machine vision where such prior knowledge *may be* available: Visual search. . . . Object recognition. . . . Recurrent organization. . . . (INTR 18)

> There are three key issues *to be considered* in developing an online palmprint identification system: Palmprint Acquisition: . . . Palmprint Feature Representation: . . . Palmprint Identification: . . . (INTR 21)

> A dependable network *should have* the ability to be changeable, i.e., reconfigurable. That is, the system should continue to run at near normal levels even when its configuration is changed due to changes in users' needs and/or system state. (INTR 44)

The results indicate that the use of explicit statements of hypothesis in CS articles is minimum. However, notable occurrences of implied hypothesis were found in the Corpus. Thus, Step b of Move 3, optional in CARS model is employed by CS in a different manner. Rather than listing research questions or hypothesis as in hard sciences such as physics, considerable usage of assumptions and implied hypothesis is found.

CONCLUSION

Since only Step a of Move 3 is obligatory for both Swales and Feak (2004) and Lewin et al. (2001, p. 56), the latter question its need and indicate the possibility of it serving a "rhetorical rather than an informational function by providing closure to the gap statement by consistently promising to fill the gap." The "purposive" element of Move 3 Step a, in the CS RA Introductions, though available as an option, is limited. On the other hand, two extremes are evident: one adhering to brevity, talking about the nature of the research in one or two sentences; the other includes those authors keen to include detailed descriptions which can be complex at times or heavily loaded with methodology. The current research undertaken by the authors could also be an extension of the previous work, or a contrast to the existing design, model, method, or technique. Use of deictic *current* is more common than *the present*. The phrase *In this paper* can occur in the beginning, middle, or end of the sentences. However, when it occurs in the beginning, it is complimented by we + verb. Having described the kind of work undertaken, in the present article, most of the authors of Computer

Science articles do not feel compelled to list research questions or hypothesis. Nevertheless, they make assumptions, sometimes. Moreover, the description of the method(s)/methodology used for the research work also needs to be accounted for if well knitted scaffolding like CARS is to be followed.

Judging on the bases of a corpus from highly reputable journals in the field of Computer science, one is likely to assume that the findings of this study may have immediate usefulness for scholars only and not for students. In my opinion, senior scholars can experiment with the genre they are working in with confidence, and sometimes come up with innovative variations which are accepted by the journal editors with less awe and criticism. In contrast, undergraduate and some graduate students need to have a model to follow and they have to make an effort to create a research space for their works. Thus, a model and guidelines based on the findings from a corpus of research papers published in esteemed journals could be used as a useful pedagogical tool. Students can be helped in not only writing introductions but also in understanding the published introductions.

Thus, this study would contribute in raising genre awareness among students and writers of scientific rhetoric. As knowing differences across important genres within their discourse community is not enough, the distinctions among the part genre of the same genre (introduction of a research article here) also need to be learned and taught to be able to be a proficient writer in any particular genre. EAP and ESP students can benefit from the findings of the present study if they are made aware of not only the rhetorical moves but also of the linguistic patterns associated with them.

APPENDIX

Example:

AS an important theme on data mining, association rule mining research [1] has progressed in various directions, . . . etc. However, there is an important form of . . . useful but could not be discovered with existing. . . . Taking stock market database as an example, . . . Suppose a database registers the price of every stock. . . . Association mining may find rules like: R1: When the prices of IBM and SUN go up, at 80 percent of probability the price of Microsoft goes up on the same day. While R1 reflects . . . , its role in price prediction is limited; and traders may be more interested in the following type of rules: R2 : If the prices of IBM and SUN go up, Microsoft's will most likely (80 percent of probability) go up the next day and then drop four days later. Unfortunately, current association rule miners cannot discover this type of rule as. . . . On the other hand, rule R2 represents . . . To distinguish these two types of transactions, we name the classical association rules as intratransaction associations and the latter as intertransaction associations. It is interesting to note that, although there is a temporal component in our

example . . . , rule R2 is different from. . . . At the first glance, the work reported by . . . [16] is similar to . . . in our examples. An episode is a sequence of events and association rules among episodes have the form PâVä) QâWä, where P and Q are event sequences and V and W are time bounds. There is however no clear concept of transaction in a sequence of events. Instead, subsequences are formed using the time bounds and episodes that occurred frequently within these subsequence are discovered using an extension of apriori algorithm [2]. These subsequences, which can contain overlapping event instances are different from transactions which share no common item instances. Even if these subsequences are treated as transactions, the rule so discovered are intratransaction, not intertransaction. Another area which seems related is the work of Bettini et al. [5], [6] which proposed the use of event structure for the discovery of temporal relationship among events in a time sequence. An event structure consists of a number of variables representing events and temporary constraints among these variables. The goal of the mining is to find temporal patterns in the sequence that could instantiate the event structure. Such work belongs to those of a belief system in which the user tries to verify their belief by providing an event structure to discover a constraint set of rules. Again, there is no clear concept of transactions here. There are many applications which can benefit from intertransaction association mining. For example, one may discover traffic jam association patterns among different highway segments: "if highway one east bound is jammed at section two, highway 12 south bound will be jammed at section five 10 minutes later." Although it is pretty challenging to find the causal structure for a complex highway system with dynamic traffic flow, intertransaction association mining may help traffic jam pattern prediction. It may also help find that the sale of a particular kind of baby toy will boost the sales of women clothes in the following week, or help discover if there is a drought in region A, there will be a flood in region B two years later from a long term regional weather database. From the above brief description, we can identify that the first contribution of our work reported in this paper is the. . . . Moreover, from the discussion in the later sections, it will be shown that our formulation can be extended to. . . . Since . . . the number of rules will increase drastically. Another contribution of our work is that. . . . The remainder of the paper is organized. . . . (INTR 49)

REFERENCES

Anthony, L. (1999). Writing research article introductions in software engineering: How accurate is a standard model? *IEEE Transactions on Professional Communication, 42*(1), pp. 38-46.

Barrass, R. (1978). *Scientists must write: A guide to better writing for scientists, engineers and students.* London: Chapman & Hall.

Blicq, R. S. (1986). *Technically-write: Communicating in a technological era* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Booth, V. (1993). *Communicating in science: Writing a scientific paper and speaking at scientific meetings* (2nd ed.). Cambridge: Cambridge University Press.

Bowker, L., & Pearson, J. (2002). *Working with specialized language. A practical guide to using corpora.* London: Routledge.

Cooper, C. (1985). Aspects of article introductions in IEEE publications. MSc dissertation, The University of Aston in Birmingham, Birmingham.

Crème, P., & Lea, M. R. (2003). *Writing at university* (2nd ed.). Maidenhead: Open University Press.

Godfrey, J. W., & Parr, G. (1960). *The technical writer: An aid to the presentation and production of technical literature.* London: Chapman & Hall Ltd.

Heidegger, M. (1954). *The question concerning technology and other essays* (pp. 3-35). W. Lovitt (Ed.). New York: Harper Torchbooks.

Hyland, K. (1998). Boosting, hedging and negotiation of academic knowledge. *Text, 18*(3), 349-382.

Kuo, C-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes, 18*(2), 121-138.

Lewin, A., Fine, J., & Young, L. (2001). *Expository discourse: A genre-based approach to social science research texts.* London/New York: Continuum.

Lindeberg, A-C. (1994). An exploratory study of knowledge claims in article introductions in three disciplines: Finance, management, and marketing. In M. Majapuro & T. Nikko (Eds.), *Talous ja kieli 11* [Business and language 11] (pp. 647-655). Helsinki: Publications of the Helsinki School of Economics and Business Administration D-2 11.

Lindeberg, A-C. (2004). *Promotion and politeness: Conflicting scholarly rhetoric in three disciplines*. Published Ph.D. dissertation, Abo Akademi University Press.

Lovitt, W. (1977). Introduction. In M. Heidegger (Ed.), *The question concerning technology and other essays* (pp. xiii-xxxix). New York: Garland.

O'Connor, M. (1991). *Writing successfully in science.* London: Chapman & Hall Ltd.

Okamura, A. (2003). *The Economic Journal of Takasaki City University of Economics, 46*(1), 49-62.

Orr, T. (1999). Genre in the field of computer science and computer engineering. *IEEE Transactions on Professional Communication, 42*(1), 32-37.

Posteguillo, S. (1995). *Genre analysis in English for computer science.* Ph.D. dissertation, Universitat De Valencia, Spain.

Posteguillo, S. (1999). The schematic structure of computer science research articles. *English for Specific Purposes, 18*(2), 139-160.

Reppen, R., & Simpson, R. (2002). Corpus linguistics. In N. Schmitt (Ed.), *Applied linguistics* (pp. 92-111). London: Arnold.

Scott, M. (1997). *WordSmith manual.* Oxford: Oxford University Press.

Shehzad, W. (2006). Computer scientists approach to "establishing a research territory." *Selected Papers from the Fifteenth International Symposium on English Teaching, 1*(1), 127-141. Crane Publishing Co. Ltd and ETA – ROC: Teipei.

Shehzad, W. (2007a). How to end an introduction in a computer science article? Corpus-based approach. In E. Fitzpatrick (Ed.), *Corpus linguistics beyond the word: Research from phrase to discourse* (pp. 227-241). New York: Rodopi.

Shehzad, W. (2007b). The explicit author in the scientific discourse. *Malaysian Journal of ELT Research, 3,* 56-73.

Shehzad, W. (2008). Move two: Establishing a niche. *IBERICA, 15,* 25-50.
Shehzad, W. (2010). Announcement of principle findings and value addition in computer science research papers. *IBERICA, 19,* 97-118.
Swales, J. M. (1981). Aspects of article introductions. Aston ESP Research Report No. *Language Studies Unit*, University of Aston in Birmingham, UK.
Swales, J. M. (1990). *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University Press.
Swales, J. M., & Feak, C. (1994). *Academic writing for graduate students*. Ann Arbor, MI: The University of Michigan Press.
Swales, J. M., & Feak, C. (2004). *Academic writing for graduate students: Essential tasks and skills* (2nd ed.). Ann Arbor, MI: The University of Michigan Press.
Swales, J. M., & Najjar, H. (1987). The writing of research article introductions. *Written Communication, 4,* 175-192.

Other Articles On Communication By This Author

Shehzad, W. (2007a). How to end an introduction in a computer science article? A corpus-based approach. In E. Fitzpatrick (Ed.), *Corpus linguistics beyond the word: Research from phrase to discourse.* Amsterdam/New York: Rodopi Publications.
Shehzad, W. (2007b). Computer scientist's approach to "establishing a research territory. Selected papers from the Fifteenth International Symposium on English Teaching, 1(1), 127-141. Crane Publishing Co. Ltd. and ETA-ROC: Teipei.
Shehzad, W. (2007c). Explicit author in the scientific discourse: A corpus-based study of the author's voice. *Malaysian Journal of ELT Research, 3,* 56-73.
Shehzad, W. (2008). Move two: "Establishing a niche." *IBERICA, Journal of the European Association of Languages for Specific Purposes, 15,* 25-50. Spain.
Shehzad, W. (2010a). Announcement of the principal findings and value addition in CS research paper introduction. *IBERICA, Journal of European Association of English for Specific Purposes, 19.* Spain
Shehzad, W. (2010b). *Pakistan Journal of Life and Social Science Genre Approaches to Comparable Discourses with a focus on Computer Science Research Article, 8*(1), 73-80.

Direct reprint requests to:

Wasima Shehzad
Yanbu University College
Kingdom of Saudi Arabia
e-mail: wasima.shehzad@yahoo.com