Feature Article

Textual Analysis of the Target Genre: IEEE Best Papers in Computer Science and Engineering

Neil H. Johnson
University of Aizu

William Rozycki
University of Aizu

Academic writing instruction can be usefully informed by corpus and textual analyses of published research in the target field (e.g. Anthony, 2001; Noguchi, 2004; Shehzad, 2007a). In this project, analysis of six award-winning computer science research articles was conducted using qualitative discourse analysis and text-mining tools. Findings confirmed earlier research suggesting that a prominent authorial voice, characterized by a high degree of metadiscourse, including wide use of personal pronouns, is an important feature of computer science research discourse (Hyland & Tse, 2004). Further, non-standard English usage was also found, raising questions for educators involved in the teaching of English for Specific Purposes (ESP). Based on results of the study, the authors offer suggestions for incorporating metadiscourse into the writing curriculum for teachers of academic research writing in computer science. The authors also discuss the notion that a pedagogical focus on writing as communication, without an undue prescriptive concern with language forms that may be very difficult to master, may also be beneficial to instructors and learners in foreign language ESP contexts.

Johnson, N. H., & Rozycki, W.. (2010). Textual analysis of the target genre: IEEE Best Papers in computer science and engineering. *OnCUE Journal, 4*(2), 80-109. *OnCUE Journal*, ISSN 1882-0220, is available at http://jaltcue-sig.org one year after publication date.

アカデミックライティング指導は、ターゲット分野で発行された研究のコーパス及びテキスト分析が有益である。(e.g. Anthony, 2001; Noguchi, 2004; Shehzad, 2007a) この研究では、6本のコンピュータサイエンスの受賞論文記事分析を、定性ディスコース分析とテキストマイニングツールを使用して行った。研究結果として、人称代名詞の広範囲利用を含め、メタディスコースの高い度合いで特徴づけられた著者の主張が、コンピュータサイエンスリサーチディスコースの重要な特徴である(Hyland & Tse, 2004)、という先行研究が立証された。さらに、非スタンダードの英語使用も見つかり、これは English for Specific Purposes (ESP) 分野の教員を含め、教育者にとって興味ある疑問を提示した。コンピュータサイエンスのアカデミックリサーチライティングの教員にとっては、ライティングカリキュラムにメタディスコースを組み入れるにはいくつかの予備的示唆がある。コミュニケーションとしてのライティングにおける教育的焦点は、習得が困難な過度の規範的懸念を除いては、外国語ESP場面における指導者と学習者両方において有益になり得るかもしれない。

Worldwide, the Institute of Electrical and Electronics Engineers (IEEE) is the largest professional body within the fields of technology, computer science, and engineering. As of 2006, a diverse and international membership of over 375,000 members (IEEE Home, 2010) was organized into over 30 separate IEEE societies in eight different regions worldwide. It is estimated that 30% of the world's research conducted in technology-based fields is published through over 100 IEEE peer-reviewed journals such as the prominent Transactions series (Wikipedia, 2010). For the undergraduate researchers in computer science at the University of Aizu, Japan, IEEE publications are a vital resource, in both content and language, for their final thesis research, upon which graduation depends.

For instructors responsible for the teaching of English for research purposes at this institute, and for English for computer science instruction in general, the IEEE research directory is an important source of information and reference on language use in the various sub-fields that comprise the computer science field. This paper derives from a concern with the guiding of computer science and engineering

students towards appropriate target research genre conventions, so that they may begin to participate in the worldwide professional discourses of research and development in their chosen fields (Flowerdew, 2001). Taking a functional perspective of language and discourse as socially constructed through use (e.g. Halliday 1985, 2004, 2006), the purpose of this research, then, is further documentation, description, and analysis of the salient ways in which English is currently used in the fields of computer science and engineering.

English and Participation in Academic Discourse

The implications of specialized and prestige forms of language have been long recognized. Bernstein (1977) described such language as "elaborated" and saw it as an important element in the creation and maintenance of social structures. Rose (2006) outlined the argument:

> *The growth of scientific knowledge has been implicated from the beginning in the evolution of modern economic systems and the global power structures they support…access to income, control and life opportunities depend on the level of technical literacy people are able to acquire through school education. (p. 237)*

Academic literacy, the ability to understand and use specialized forms of language, is an important element in the creation and maintenance of social hierarchy. The role of English as the central medium worldwide for the transmission and communication of science and engineering research, through such publications as the IEEE Transactions journals, is controversial and an important issue for educators concerned with the teaching of English for Specific and Academic Purposes (Flowerdew, 2001; Burrough-Boenisch, 2003; Cho, 2009). Flowerdew (2001) analyzed the problems that non-native speakers of English (NNS) have in seeking to publish their research through a survey of applied linguistics related journal editors. The reported findings suggested

that publication was problematic for NNS authors, and issues were associated with "surface errors, parochialism, absence of authorial voice and nativized varieties of English" (Flowerdew, 2001, p. 121). It is worth noting how negative assessments of this research tends not to be focused on the content of the research itself, but largely on discourse community conventions, defined as underlying values and goals for the writing community that influence various aspects of text production, such as the use of authorial voice.

Flowerdew's (2001) findings suggest that non native-speakers of English have a disadvantage in competitive research publication that is a central aspect of research and career advancement in any academic field. This conclusion was echoed in Cho's (2009) assessment of the issues facing both Korean graduate students and professors in the field of science and engineering, reporting that, "About one-third of the respondents felt the referees and editors were biased" (p. 237). In the field of biology (Burrough-Benisch, 2003), however, inquiry into reviewers of published research found evidence for a different situation: non-native specialists reviewing journal submissions who were more likely to evaluate content, rather than to criticize the language use. Burrough-Benisch concluded that "NNS [non-native speakers] were not at a linguistic disadvantage" (p. 235) in the discourse community of biology research. The reason for this contradiction might be uneven increases in the numbers of non-native speakers in positions of academic authority, something Burrough-Benisch (2003) has also surmised.

Global English and the Native Speaker

The global spread and usage of English as the language of research in such fields as computer science has been a prominent subject in applied linguistics research for well over twenty years (see Kachru, 1986), and the ideological and political aspects of this have been widely discussed (Phillipson, 1992; Pennycook, 1994; Widdowson,

2003; Mair, 2003; Seidlehofer; 2001; Kachru, 2005). It has been argued however, that this awareness of English as a language used by NNS has yet to be translated into meaningful changes in the teaching of English, which as Seidlhofer (2001) argued, (following Pennycook, 1994; Medgyes, 1994) is still largely dominated by the concept of the native-speaker (NS):

> *What constitutes a valid target is still determined with virtually exclusive reference to native-speaker norms. True, at least the perception of what constitutes 'native speakers' is widening, but a question in urgent need of exploration is just what the 'English' is that is being taught and learnt in this emerging global era, how it squares with the sociopolitical and socioeconomic concerns discussed in the profession, and what its relevance is for the subject taught in classrooms all over the world. (p. 135)*

Seidlhofer (2001) suggested that the term "English as a lingua franca" should be used to conceptualize the use of English by the majority of its speakers, "those who primarily learnt English as a lingua franca for communicating with other lingua franca speakers" (p.139). The key point of this shift in language use is that users no longer concern themselves with native-speaker models of use in their own discourse communities, but become centrally concerned with "efficiency, relevance and economy in language learning and language use" (Seidlhofer, 2001, p.141). Kachru and Smith (2009), however, warned against the use of terms such as global English or lingua franca, claiming that these terms themselves mask the reality of language variation. They called for more familiarity with variation itself across domains of language use, as a way of cultivating and acknowledging the linguistic reality of language as it adapts to new contexts of use (p. 2).

The issue of standard and non-standard forms of language and what should be taught and accepted in the language classroom, and for publication, is complex. If prescriptive "NS norms" are no longer the models for usage, does this mean that variations on standard forms are

always acceptable? How can we distinguish between local variation and simple error? Is there a danger in setting aside NS norms that educators could unwittingly set up learners to be further disadvantaged through an inability to conform to what may be seen as international academic standards?

There are no easy answers to these questions, and theorists (see Rubdi & Saraceni, 2006) working within the World Englishes and English as an International Language frameworks have been constructing a position that rejects prescriptive models of native speaker language correctness, yet accepts that standards that allow for clear and precise communication are in the interests of all language users, regardless of their goals and their uses of the language. Indeed, Gupta (2006) suggested that "usage" is the appropriate criteria for establishing the parameters of what is acceptable. The suggestion that usage provides instructors with guidance for the classroom has been important in establishing corpus linguistics and discourse analysis as tools to inform pedagogical practice. In the following section, we highlight some ways in which corpus analysis has provided insights into the development of the language of computer science.

Developments in Science Discourse

Corpus-based research into personal pronoun use in scientific writing highlights some recent shifts in the way in which science research in general is being communicated in language. Kuo (1999) reported on changes in pronoun use over time, in keeping with Halliday's (2004) view of discourse as a developmental system. He suggested that, after the nineteenth century, a move towards depersonalized registers of science created the effect of objectivity in which the scientists themselves were effectively erased from the discourse. Such impersonal style was commonly viewed as "demonstrating a grasp of scholarly persuasion as it allows the research to speak directly to the reader" (Hyland, 2001, p. 208). There are a number of rhetorical strategies available for writers

to represent their authorial self within a text, ranging from the strongly visible to a less visible presence. According to Hyland (2001), one of the ways to do that is the use of personal pronouns, which allows authors to present information both subjectively (e.g. *we investigate*, *we suggest*) and interpersonally (e.g., *as we can see*). Kuo (1999) claimed that in the latter half of the 20th century more dynamic and flexible styles were increasingly developing in different disciplines, equating this with increasing professionalism:

> *The scientist must claim the significance and contribution of his/her research to the discipline on the one hand, but appeal modestly to both editors and expected readers – his/her peer researchers– seeking their approval and acceptance on the other. (p. 122)*

He concluded that writers in the sciences use first-person plural pronouns frequently, particularly an inclusive *we*, to refer to the discipline as a whole (p. 136). Chang and Swales (1999) characterized this shift as including "more personal comment, narration and stylistic variation." Elsewhere, this has been described as an increase in "writer visibility" (Kaplan et al., 1994). The function of pronoun use in this context allows authors "to emphasize their personal contributions to their field of research and how to seek cooperation and stress solidarity with expected readers and their disciplines" (Kuo, 1999, p. 136).

To further investigate the use of authorial voice in scientific discourse, Shehzad (2007b) used a corpus of 56 computer science research articles published in the IEEE Transactions series. She found that the use of personal pronouns was common and added to a more personal style through the creation of an authorial voice that she described as "explicit, firm and assertive" (p. 68). Shehzad attributed this style to the "discipline's newness" and the rapid innovation and development of research that necessitates a prominent authorial presence within the text. Shehzad (2007b) further distinguished between exclusive and inclusive use of "we" in which exclusive use refers to the writer or

writers, and inclusive use of "we" references the writer and the reader together. In both Hyland (1999) and Shehzad (2007b) the exclusive "we" was noted to be significant in science research discourse. In Shehzad's corpus, over 90% of the cases of the pronoun "we" were "exclusive," referring to the authors themselves.

The use of personal pronouns is the most obvious way in which authors can insert themselves into a text. There are however other possibilities for authorial presence, including the use of "metadiscourse." The term metadiscourse is broadly defined by Adel (2006) as language in a text that functions to both offer both "guidance and interaction" (p. 43) for and between the reader and the author. Hyland and Tse (2004) suggested that metadiscourse includes "the 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). Hyland and Tse offered a broad model for understanding the interactive nature of writing, including pronoun use as authorial self-reference, and suggest that metadiscourse as an essential part of the understanding of propositional meaning, cannot be separated from the central meaning of the text. In other words, metadiscourse is not simply the language that binds the elements of textual meaning together, but it is a central part of how a text becomes to mean what it does. Other researchers (Mauranen, 1993; Valero-Garces, 1996) concerned with meta-language in academic text have used a rather narrower frame that refers to language about the text itself, where the author may signal intentions or describe in summary what has been accomplished in a particular section. This narrower definition is primarily concerned with textual cohesion and signaling relationships between parts of the text. As such, it excludes rhetorical questions, hedging, boosting, pronoun use and reference outside the text itself, which are included in Hyland and Tse's later (2004) model.

While we acknowledge that the narrower approach to metadiscourse is problematic in some ways (see Hyland & Tse, 2004), since other

aspects of the text such as hedging are also clearly a part of metatextual discourse, in this paper we adopt the narrow approach for the purposes of cross-disciplinary comparison with earlier research findings using this model. Mauranen (1993) completed one of the earliest studies of metadiscourse in academic text and differentiated between and compared "Anglo-American," or native speakers of English, and Finnish authors and found that there were significant differences in the amount of metatext used in economics journal articles written by these different authors. The data resulted from a comparison of two pairs of texts chosen for their similarity in terms of genre (academic research reports), field, (economics) and topic (forestry economics and taxation). In both pairs, one text was written in English by a Finnish economist, the other by an economist who was a native speaker of English. Mauranen (1993) concluded that the "Anglo-American rhetorical style" shows more concern for the reader, who is generally more explicitly oriented and guided throughout the text.

Valero-Garces (1996) followed Mauranen's (1993) research design and distinguished between economic research texts written in English by "Spanish-speaking academics" from mainland Spain and by "Anglo-American academics" in order to further explore cultural differences in academic rhetoric. Valero-Garces (1996) defined the narrow approach to metadiscourse according to the following four textual functions and provided examples from the corpus for each type:

1. Connectors. Basically conjunctions, and adverbial and prepositional phases, which indicate relationships between the text itself and its content, e.g., "however," "for example," "therefore."
2. Reviews or earlier markers. These contain an explicit indicator that an earlier stage of the text is being repeated or summarized e.g., "so far we have assumed that," "as previously mentioned."
3. Previews or later markers. These contain an explicit indicator that a later stage of the text is being anticipated. They can

be words, expressions, phrases and sometimes clauses or abbreviated clauses, e.g., "we show below that..."

4. Action markers. These are indicators of discourse acts performed in the text, e.g., "the explanation is," "to illustrate this" (p. 284).

The results suggested that the "Anglo-American" writers used significantly more metatext than the "Spanish-speaking" writers (Table 1). Valero-Garces (1996) concluded that Anglo writers (ET1 and ET2) seemed to be more concerned with guiding and orienting the reader than the Spanish-speaking writers (ST1 and ST2), and showed more explicit presence of the writer in the text. The Spanish-speaking writers put more emphasis on propositional content and seemed to prefer a more impersonal and implicit style of writing. The findings are presented here as a percentage of examples per sentence, so that direct comparisons may be made with the computer science discipline. There is also evidence, in the higher numbers of connectors and action markers in particular, that the Anglo-American writers show more concern with guiding and orienting the reader than the Spanish-speaking writers, and, it is suggested, generally show a "more explicit presence of the writer" (p.269) within the text.

Table 1. Metatext in Spanish-English Economics Texts (Valero-Garces, 1996)

	ST1		ET1		ST2		ET2	
Metatext	*N*	%	*N*	%	*N*	%	*N*	%
Connectors	62	20.5	60	28.4	23	12.7	29	17.6
Reviews	3	0.9	25	11.8	6	3.3	9	5.4
Previews	7	2.3	20	9.4	4	2.2	9	5.4
Action markers	21	6.9	35	16.5	24	13.3	24	13.5

Notes. ST = Spanish text; ET = English text; % = metatext per sentence.

Rhetorical differences in regard to authorial presence according to academic disciplinary culture have also been explored. Dahl (2004) took a contrastive corpus analysis approach and investigated writer presence in three languages (English, French and Norwegian) and across three disciplines (economics, linguistics and medicine) in order to investigate if academic discipline or language is the most important variable governing the pattern of metadiscourse in academic text. The corpus consisted of 180 refereed research articles within those languages and disciplines. The analysis suggested that the language variable was the most important within economics and linguistics, where English and Norwegian texts showed very similar patterns and used much more metatext than the French text. In contrast, within the field of medicine all three languages displayed a uniform pattern of infrequent metadiscourse.

Although Shehzad (2007b) found wide use of the pronoun "we" in computer science discourse, to date, there has been little research done on the use of metadiscourse in computer science research writing. Hyland and Tse (2004), however, compared the broader use of metadiscourse across disciplines, including computer science, for doctoral and masters thesis work. Their findings were instructive: "Computer science tended to differ from this general picture of impersonality in scientific discourse, displaying relatively high frequencies of self-mentions and engagement markers" (p.176). The present study provided a further step towards understanding the use of this important feature in computer science research discourse and we examined research published in the field. Further, by concentrating in detail on a smaller corpus, we were able to look at variation within the field, which is characterized by contributions from authors from many different backgrounds. At the outset, the research questions guiding this research were:

1. What are the distinguishing features of computer science discourse that may inform the teaching of research-based

writing in this field?

2. If there are distinguishing features, how are they achieved lexically?

Methods

In this study, six recently awarded Best Papers from the *IEEE Transactions for Computer Science and Engineering Journal* were selected and analyzed for recurring features of language style and use. The criteria for selection for this corpus were that the paper be awarded "Best Paper" by the IEEE society, fall within the field of computer science, and be recently published (2006-2008). From the possible papers available, six papers (Table 2) were randomly selected to provide a representative sample. One further criterion was article length. We wanted each paper to be a representative research paper in this field; since computer science papers are typically between 10 and 25 pages in length, one sample paper was rejected for being too short, at only 3 pages in length.

The analysis process started with both researchers reading the texts with a view to noticing recurrent features that would form the basis of a more detailed analysis. Once the most salient features were agreed upon, further token counting could proceed using, in the case of personal pronouns, a text- mining tool. The pronoun token counts in this work were performed with software developed by Schmitt and Christianson (1998), called UNEIM, based on the sentence-finder and text-mining techniques described in subsequent work of the developers (Schmitt & Christianson, 2006). Text-mining, also known as text data mining, is defined as "the process of extracting interesting and non-trivial patterns or knowledge from unstructured text documents" (Tan, 1999, p. 1). In our case, we were interested in the size of each text and set the software to analyze each text according to the number of sentences. The software was programmed to create further analysis of each text in terms of sentences in which pronouns were found, and

Table 2. Overview of Selected Papers

Authors	Title	Publication	Affiliation	Background
1. Chou & Miao	Rate-Distortion Optimized Streaming of Packetized Media	*IEEE Transactions On Multimedia*, 8, (2). (2006).	University of Washington, Seattle. Sony Research Lab, Santa Clara, CA.	USA and China
2. Tsividis	Mixed-domain systems and signal processing based on input decomposition.	*IEEE Transactions on Circuits and Systems*. 53, (10). (2006).	Columbia University, NY.	USA
3. Panigada & Galton	Digital background correction of harmonic distortion in pipelined ADCs.	*IEEE Transactions on Circuits and Systems*. 53, (9). (2006)	University of California, San Diego.	USA and Italy
4. Heydari & Pedram	Model-order reduction using variational balanced truncation with spectral shaping.	*IEEE Transactions on Circuits and Systems*. 53, (4). (2006).	University of Southern California. University of California, Irvine.	Iran and USA
5. Chen, et al.	Integration of digital stabilizer with video codec for digital video cameras.	*IEEE Transactions on Circuits and Systems for Video Technology*. 17, (7). (2007).	National Taiwan University. Novatek Inc., Taiwan. MIT, USA.	Taiwan
6. Bletsas, et al.	Cooperative Communications with outage-optimal opportunistic relaying.	*IEEE Transactions on Wireless Communications, 6(9)*. (2007).	Aristotle University of Thessaloniki, Greece. Kyung Hee University, Korea. MIT, USA.	Greece, S. Korea, and USA

to list collocations for each pronoun usage. The other aspects of the texts of interest to the researchers were analyzed holistically by careful notation, collection, and counting of samples using established criteria, which were then checked using the freeware corpus linguistics tool AntConc (Anthony, 2005). This corpus tool allowed for clarification

Table 3. Overview of Analysis of the Six Best Papers of IEEE

Authors	Number of sentences	Pronoun Use: "We" Exclusive	Pronoun Use: "We" Inclusive	Metatext Words Or Phrases	Metatext % Per Sentence	Variational English Examples
1. Chou & Miao	541	87	3	208	38.4	0
2. Tsividis	556	29	0	214	38.4	0
3. Panigada & Galton	329	0	0	146	44.3	0
4. Heydari & Pedram	414	14	1	114	27.5	31
5. Chen, et al.	467	29	3	101	21.6	14
6. Bletsas, et al.	323	47	0	99	30.6	0

of the items noticed in the text, listing of common collocations, and ensured an accurate token count.

Following Shehzad (2007b), we sorted pronoun use according to an exclusive or inclusive function. Similar to Mauranen (1993) and Valero-Garces (1996), we were concerned with the "narrow" definition of textual metadiscourse:

1. Connectors
2. Reviews or earlier markers.
3. Previews or later markers.
4. Action markers.

The results are presented as a percentage of use within the total number of sentences, to allow for direct comparison with the texts in Valero-Garces' (1996) study. In this study, markers such as "let p equal…" were excluded on the grounds that they were generically determined and offered no real rhetorical choice. For the purposes of direct comparison, we therefore excluded such general markers, also.

Table 4. Textual Examples: Use of Pronoun "We"

Authors	Examples Of Pronoun Use
1. Chou & Miao	In this paper, we show, for arbitrary encodings and packetizations We present the major ideas in our paper as follows
2. Tsividis	In this section we present the principle that will form the basis We consider a continuous-time linear
3. Panigada & Galton	N/A
4. Heydari & Pedram	Now, we proceed with describing the new procedure We still need to solve the Lyapunov equations to obtain
5. Chen, et al.	We begin this section by giving an overview of the existing digital stabilization techniques To answer the question, we set up an experiment
6. Bletsas, et al.	In this paper, we present single-selection - opportunistic- relaying We propose simple opportunistic relaying schemes

Results

Through the process of reading the texts closely, three features were agreed upon as salient: (1) use of the personal pronoun "we," (2) an explicit and prominent authorial voice, and (3) variational, or non-standard use of English. Prominent authorial voice was defined as an example in the text where either the author or authors signal their authorial intention to the reader, in a meta-description of the progress of the research article. Following Valero-Garces (1996), we first provided an overview of metatext in all of the sample articles (Table 3) and then provided examples of pronoun "we" usage from each paper (Table 4). Variational English was defined as language use that exhibited variation

Table 5. Textual Analysis: Metatext

Authors	Connectives		Reviews		Previews		Action markers	
	N	%	*N*	%	*N*	%	*N*	%
1. Chou & Miao	142	26.2	9	1.6	22	4.0	35	6.4
2. Tsividis	127	22.8	27	4.8	24	4.3	36	6.5
3. Panigada & Galton	100	30.3	19	5.7	11	3.3	16	4.8
4. Heydari & Pedram	83	20.0	9	2.7	10	4.1	12	2.8
5. Chen, et al	64	13.7	9	1.9	14	1.9	14	2.9
6. Bletsas, et al	74	22.9	5	1.5	8	2.4	12	3.7

from language forms described in formal descriptions of English syntax (e.g., Quirk, Greeenbaum, Leech, & Svartvik, 1985) and/or the stylistic conventions of the discourse genre.

Pronoun Use and Authorial Presence

The use of the exclusive personal pronoun "we" was found in five out of six papers in the corpus and was prominent in papers 1 and 6 especially, with 90 and 47 tokens respectively. As in Shehzad's (2007b) study, the use of inclusive pronoun "we," referring to the writers and the reader, was hardly found; only 3% of the cases were inclusive. Examples of usage from each paper are presented in Table 4.

In Table 5 we provide an overview of the use of metatext in this corpus, according to the model developed by Mauranen (1993) and Valero-Garces (1996).

Table 6. Textual Examples: Variational English Use

Authors	Textual Examples
1. Chou & Miao	N/A
2. Tsividis	N/A
3. Panigada & Galton	N/A
4. Heydari & Pedram	The new variational SBT attains, in average, 30% As the future work, we will investigate Interconnect technology parameters are varying substantially The matrix is partitioned into to two submatrices …the Safanov's algorithm is used
5. Chen, et al.	This type of approaches is more economic for digital The motion field of an image sequence is contributed by three kinds of motion This increment due to larger input frame for stabilizer
6. Bletsas, et al.	N/A

Variational English

The presence of variational English in two papers written by NNS in our corpus is also noteworthy. Examples from each of the papers that include this language are presented in Table 6. Analysis of the non-standard English in this sample suggests that the variations lie in two main areas. First, there may be examples of language use that do not conform to the genre-based norms of the discourse, and though grammatically

correct, may violate expectations of register and be conversational in tone. An example of this from paper 5 is: *A bigger image margin means a larger camera shaking can be handled*. In this sentence, the idea that the authors wish to communicate is not obfuscated by the variational use of *larger camera shaking* to describe a more stable camera product. The use of *bigger*, very much conversational in tone, also successfully communicates the intended idea.

The second area of variational use can be described as syntactic. There are, in these sample texts, several instances of variations with articles of speech such as "the" and "an." For example, the article may be missing: *solution of Lypanov equation*. An article may be used where not required, as in paper 4: *the Safanov's algorithm is used*. It is important to note that a missing article is unlikely to impede effective communication, nor is the addition of a redundant article likely to limit understanding, and it may well be for this reason that such examples are being found in these example texts.

Discussion

In this section we discuss the findings from the IEEE award winning papers, recently published in the field, in terms of pronoun use, metadiscourse and thirdly, variational English. With regard to pronoun use, the findings here confirm the research conducted by Shehzed (2007b) into explicit authorial presence in computer science discourse. Although there was considerable variation within the corpus, there was a general trend of using "we" to explicitly self-reference the authors within the texts. Interestingly, use of exclusive "we" is also found in paper 2, which has a single author, perhaps as recognition of the collaborative nature of much research in the field.

In terms of metadiscourse, comparisons can be made with the economics texts in the Valero-Garces (1996) study. In this early study, considerable variation was found in the use of metatext according to

the cultural and rhetorical backgrounds of the authors. For example, the first "Anglo-American" author used connectors in 28.4% of the sentences written. The second Spanish-speaking author used them in 12.7% of sentences. This finding is reflected in the present study, where there is considerable variation within the corpus. Panigada and Galton (text 3 in the sample) with American and Italian backgrounds, used connectives in 30% of their sentences, compared with Chen et al. (paper 5), with Taiwanese backgrounds, who used connective devices in 13.5 % of their sentences. This variation is perhaps a reflection of the diversity of rhetorical training and approach among researchers who currently comprise the field.

Comparing across disciplines, computer science metadiscourse is generally within the same range of those writing in economics, which is a social science. For example, all but one of the computer science texts analyzed (paper 5) have more connectors as a percentage per sentence than the second Anglo-American text (17.6%) in the Valero-Garces (1996) study. The use of reviews and previews was generally similar to the economic texts (between 2% and 5%). However, the use of action markers in the Valero-Garces corpus (16.5% and 13.5% for the Anglo-American texts), was notably more than that used in the computer science texts, where 6.5% was the highest usage recorded (in text 1).

This general trend, i.e., a level of metadiscourse similar to the high level of "Anglo-American" authors in economics, reflects the findings of Hyland and Tse (2004), who found that computer science was more personal and interactive than the harder sciences such as biology, but less than other subjects such as business studies. Hyland and Tse speculated that the reason for the marked difference between computer science and the harder sciences of biology and electrical engineering was that "research in computer science tends more to the everyday world and as a result its metadiscourse has evolved" (p. 176).

Also, in view of the corpus data in the current study it may be

worth reflecting on an observation made by Ramanathan and Atkinson (1999) that Asian cultures may find difficulty in presenting a clear and "unique" sense of self in research writing. The authors of paper 1 have an American and Chinese background and yet the frequent use of the personal pronoun "we" (87 items), reviews, previews, and action markers is evidence that these authors were able to negotiate their own authorial place within their research article.

Indeed, the signaling of authorial intent, using the personal pronoun or other rhetorical strategies, was evident to some degree across all of the samples in this small corpus. This signalling provides us with perhaps the most concrete suggestion for teaching undergraduate thesis writing in Japan, the initial impetus for the study. Novice writers can be familiarized first with the notion that research writing involves interaction with a reader, whose needs must be anticipated and accommodated. With this understanding, the concept of textual cohesion in a paper will make more sense and subsequently learners can be encouraged to discover techniques for achieving cohesion within their own work. For example, this could be achieved through analyzing exemplars from within their own academic fields (Cheng, 2008). A corpus of different textual examples that learners themselves collect can be used to highlight the different ways that writers can guide their reader. Table 7, with data from the present study, provides some initial guidance for teaching the narrow version of metadiscourse, which is primarily concerned with textual cohesion.

Hyland and Tse (2004) warned against treating metadiscourse at the "surface level" in dealing only with this feature as a stylistic element of text. Indeed, novice writers do need to become aware of the essentially communicative and interactive nature of research writing as an important step to understanding the more nuanced version of metadiscourse outlined in the Hyland and Tse (2004) model. With undergraduate writers, however, it may be productive to start with a narrower textual cohesion based approach. Further work could then

build on this, to approach the more interactive and interpersonal aspects of metadiscourse. Concrete suggestions for how each of these functions can be deployed in the text and how this serves the needs of the reader might be a useful place to start with undergraduate writers. Commonly used lexical items can then be described and illustrated.

Table 7. Most Frequently Used Lexical Items for Functions

Rank	Connectives	Reviews	Previews	Action Markers
1	However	As described above	In section… we	We show/it shows
2	Since	Previous section	In this section	Is/was/were presented
3	Therefore	Mentioned above	In the next section	We present
4	For example	Above discussion	Next	In summary
5	Although	Mentioned earlier	Later	We develop

These lexical examples, once further developed and presented within a framework of textual cohesion, provide a potentially useful addition to the teaching repertoire for English for computer science research, because there is growing evidence that this discourse community has developed a preference for clear signals for the reader to follow. This development may be a result of the position of the discipline between science and everyday concerns and may also be due to the increasingly diverse participants of this field, since clear and direct signaling allows the reader to understand the structure and organization of what might be otherwise opaque research documentation.

The presence of variational English in selected Best Papers in computer science—published research that has passed not only the

traditional "gatekeepers" of editors and reviewers, but also the selection committees that decide the prestige of Best Paper awards—indicates a gradual change in the discourse of international science. It is reasonable to assume that currently a majority of scientists and engineers are NNS, judging by the fact that roughly half of all scientists and engineers within the United States itself are NNS (National Science Foundation, 2003). Kourilova (1998) studied the interaction between peer reviewers of science articles and NNS writers and argued, "if they are to produce discourse that would not violate the native speaker's expectations, they have to master not only linguistic but also socio-cultural strategies" (p. 113). This argument presumes, of course, that the "native speaker's expectations" are the standard to aim for, and that it is a native speaker who is reviewing and ultimately reading the paper. As the present findings suggest, and as Burrough-Benisch (2003) has also argued, in science fields this may simply not be the case. The results of this research indicate that, contrary to Flowerdew (2001) and Cho's (2009) earlier suggestions, NNS research authors are not being disadvantaged because of language related issues, at least in the field of computer science.

On the other hand, prescriptivist grammarians, and some English instructors, may very likely consider the variational use found in two of the papers reviewed in the current study as simply unacceptable instances of mistakes in English that require correction. While in one sense it is clearly true that there is variation from standard form, we would argue that the fact that these papers were named Best Papers within their own fields, is concrete evidence that variational language use does not interfere with the communication of the research content. Previous work on the analysis of NNS written texts (Hinkel, 2002) and their difference with NS texts has taken the position that understanding the differences can be used as a starting point to remedy what may be seen as deficiencies in the NNS writing performance. Gupta (2006) however, makes the following pedagogical suggestion that offers a way of balancing concerns with clear and appropriate language instruction

and understanding of the ways in which English is being used as a tool for international communication: "Teachers should do their best to establish what they should correct firmly, what they should correct tentatively, and what they should accept as correct. It is only through close attention to usage that this can be established" (p. 107).

Conclusion

In this paper, we examined science writing metadiscourse, narrowly defined as language describing and referring to the text itself, in computer science research articles and compared it with previous research using the same model for economics research (Valero-Garces, 1996). We found that for this definition of metadiscourse, computer science generally included a high frequency of text that was concerned not with the propositional meaning of the research, but with guiding the reader and signaling authorial intentions and accomplishments within the text. We also found wide variation within computer science research in the way that authors negotiate this rhetorical function. Our findings confirm earlier research from Shehzad (2007b) and Hyland and Tse (2004) that suggested the computer science research community seems to prefer a prominent and interpersonal style of signaling within the research article.

Research into metadiscourse from a broader, more interpersonal perspective (Hyland & Tse, 2004) has established its importance as the interface between an author and the reader, and as an integral part of propositional claims within the research report. These preliminary findings here confirm the importance that the interactive elements of research writing have in the establishment of organizing research claims, and we suggest that instruction in academic writing for research purposes needs to reflect this importance. For beginning researchers in particular, learning how to place oneself into a text appropriately, and using this voice to accommodate the requirements of the reader, will continue to be challenging. In future research, we intend to apply

the broader, more interactive model of metadiscourse, developed by Hyland and Tse (2004), to a larger corpus of computer science research writing and use these findings to further develop English instructional materials for teaching undergraduate researchers in computer science.

In countries such as Japan, China, or Korea, where a local variety of English does not have an independent basis, there may still be a need to adopt "exonormative" forms (Kirkpatrick, 2007), but influences from the native language, as long as these do not interfere with intelligibility, can still find expression. Canagarajah (1996) and Flowerdew (2001, 2008) have argued that intelligibility to a scholarly community, rather than standard English, should be the criteria for what is and what isn't acceptable English. To fully understand this controversial issue, and to clarify what the implications might be for ESP instruction, further consultation with reviewers and editors of science journals is necessary. We close with the following provocative quote, which perhaps sums up the situation as it has evolved today, to provide the context for further discussion and research:

> *There is no justification for doggedly persisting in referring to an item as an error if the vast majority of the world's L2 English speakers produce and understand it. Instead, it is for L1 speakers to move their own receptive goal posts and adjust their own expectations as far as international (but not intranational) uses of English are concerned…This also drastically simplifies the pedagogic task by removing from the syllabus many time consuming items which are either unteachable or irrelevant for EIL [English as an International Language].*
>
> (Jenkins, 2000, p.160).

Acknowledgments

The authors would like to thank the *OnCue Journal* editors for helpful comments on an earlier version of this paper.

References

Adel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.

Anthony, L. (2001). Characteristic features of research article titles in computer science. *IEEE Transactions on Professional Communication, 44*(3), 187-194.

Anthony, L. (2005). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. In *Proceedings of IWLeL 2004: An Interactive Workshop on Language e-Learning* (pp. 7- 13). Limerick, Ireland: IEEE.

Bernstein, B. (1977). *Class, codes and control (Volume 3)*. London: Routledge and Keegan Paul.

Bletsas, A., Shin, A., & Win, M. Z. (2007). Cooperative communications with outage-optimal opportunistic relaying. *IEEE Transactions on Wireless Communications, 6*(9), 3450-3460.

Burrough-Boenisch, J. (2003). Shapers of published NNS research articles. *Journal of Second Language Writing, 12*(3), 223-243.

Canagarajah, A. S. (1996). "Nondiscursive" requirements in academic publishing, material resources of periphery scholars, and the politics of knowledge production. *Written Communication, 13,* 435–472.

Chang, Y.-Y., & Swales, J. M. (1999). Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers? In C. N. Candlin, & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 145–167). Harlow: Longman.

Cheng, A. (2008). Analyzing genre exemplars in preparation for writing: The case of an L2 graduate student in the ESP genre-based instructional framework of academic literacy. *Applied Linguistics, 29*(1), 50-71.

Cho, D. W. (2009). Science journal paper writing in an EFL context: The case of Korea. *English for Specific Purposes, 28*(4), 230-239.

Chou, P.A., & Miao, Z. (2006). Rate-distortion optimized streaming

of packetized media. *IEEE Transactions on Multimedia, 8*(2), 390-404.

Dahl, T. (2004). Textual metadiscourse in research articles: a marker of national culture or of academic discipline? *Journal of Pragmatics, 36*(10), 1807–1825.

Flowerdew, J. (2001). Attitudes of journal editors to nonnative speaker contributions. *TESOL Quarterly, 35*(1), 121-150.

Flowerdew, J. (2008). Scholarly writers who use English as an Additional Language: What can Goffman's "Stigma" tell us? *Journal of English for Academic Purposes, 7*(2), 77-86.

Gupta, A. F. (2006). Standard English in the World. In R. Rubdy & M. Saraceni (Eds.), *English in the world: Global rules, global roles* (pp. 95-109). London: Continuum.

Halliday, M. A. K. (1985). *Spoken and written language*. Geelong, Australia: Deakin University Press.

Halliday, M. A. K. (2004). The grammatical construction of scientific knowledge: The framing of the English clause. In J. J. Webster (Ed), *The language of science* (pp. 102 - 134). London: Continuum.

Halliday, M. A. K. (2006). Written language, standard language, global language. In B. B. Kachru, Y. Kachru, & C. Nelson (Eds.), *The handbook of world Englishes* (pp. 349-365). London: Wiley-Blackwell.

Heydari, P., & Pedram, M. (2006). Model-order reduction using variational balanced truncation with spectral shaping. *IEEE Transactions on Circuits and Systems, 53*(4), 879-891.

Hinkel, E. (2002). *Second language writers' text: Linguistic and rhetorical features.* London: Lawrence Erlbaum.

Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. In C. N. Candlin, & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 99-121). London: Longman.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes, 20*(3), 207-226.

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

IEEE Home. (2010). IEEE Organization. Retrieved June 1, 2010, from http://www.ieee.org/index.html

Jenkins, J. (2000). *The phonology of English as an international language.* Oxford: Oxford University Press.

Kachru, B. (1985). Standards, codification, and linguistic realism: The English language in the Outer Circle. In R. Quirk, & H. G. Widdowson (Eds.), *English in the world: Teaching and learning the language and literature* (pp.11-30). Cambridge: Cambridge University Press.

Kachru, B. (1986). *The alchemy of English: The spread, functions and models of nonnative Englishes*. Oxford: Pergamon.

Kachru, Y. (2005). Teaching and learning of world Englishes. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 155–174). Mahwah, NJ: Laurence Erlbaum.

Kachru, Y. & Smith, L. (2009). The Karmic cycle of world Englishes: some futuristic constructs. *World Englishes, 28*(1), 1–14.

Kaplan, R., Cantor, S., Hagstrom, C., Kammi-Stein, L. D., Shiotani, Y. & Zimmerman, C. B. (1994). On abstract writing, *Text, 14*(3), 401-426.

Kirkpatrick, A. (2007). *World Englishes: Implications for international communication and English language teaching*. Cambridge: Cambridge University Press.

Kourilova, M. (1998). Communicative characteristics of reviews of scientific papers written by non-native users of English. *Endocrine Regulations, 32*(1), 107-114.

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

Mair, C. (Ed.). (2003). *The politics of English as a world language*. Amsterdam: Rodopi.

Mauranen, A. (1993). Contrastive ESP rhetoric: Metatext in Finnish-English economics texts. *English for Specific Purposes, 12*(1), 3-22.

Medgyes, P. (1994). *The non-native teacher*. London: Macmillan.

National Science Foundation. (2003). *The science and engineering workforce: Realizing America's potential*. Retrieved April 10, 2010, from http://www.nsf.gov/nsb/documents/2003/nsb0369/nsb0369.pdf

Noguchi, J. (2004). A genre analysis and mini-corpora approach to support professional writing by nonnative English speakers. *English Corpus Studies, 11*(1), 101-110.

Panigada, A., & Galton, I. (2006). Digital background correction of harmonic distortion in pipelined ADCs. *IEEE Transactions on Circuits and Systems, 53*(9), 1885-1895.

Pennycook, A. (1994). *The cultural politics of English as an international language*. London: Longman.

Phillipson, R. (1992). *Linguistic imperialism*. Oxford: Oxford University Press.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of English*. London: Longman.

Ramanathan, V., & Atkinson, D. (1999). Individualism, academic writing, and ESL writers. *Journal of Second Language Writing, 8*(1), 45-75.

Rose, D. (2006). Science discourse and industrial hierarchy. In J. R. Martin, & R. Veel (Eds.), *Reading science: Critical and functional perspectives on discourses of science* (pp. 236-265). London: Routledge.

Rubdi, R. and Saraceni, M. (2006). *English in the world: Global rules, global roles*. London: Continuum.

Schmitt, L. M. & Christianson. K. T. (1998). Pedagogical aspects of a UNIX-based network management system for English instruction. *System, 26*(4), 567-589.

Schmitt, L. M., Christianson, K. T., & Gupta, R. (2006). Linguistic computing with UNIX tools. In A. Kao, & S. R. Poteet (Eds.), *Natural language processing and text mining* (pp. 227-268). London: Springer.

Seidlhofer, B. (2001). A description of English as a lingua franca. *International Journal of Applied Linguistics, 11*(2), 2001.

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* (pp. 243-255). Amsterdam: Rodopi.

Shehzad, W. (2007b). Explicit author in scientific discourse: A corpus-based study of the author's voice. *Malaysian Journal of ELT Research, 3*, 56-73.

Tan, A. -H. (1999). Text mining: The state of the art and the challenges. In *Proceedings of PAKDD workshop on knowledge discovery from advanced databases* (pp. 65–70). Beijing, China.

Tsividis, Y. (2006). Mixed-domain systems and signal processing based on input decomposition. *IEEE Transactions on Circuits and Systems, 53*(10), 2145-2156.

Valero-Garces, C. (1996). Contrastive ESP rhetoric: Metatext in Spanish-English economics texts. *English for Specific Purposes, 15*(4), 279-294.

Widdowson, H. (2003). *Defining issues in English Language teaching*. Oxford: Oxford University Press.

Wikipedia. (2010). Institute of Electrical and Electronics Engineers. Retrieved June 25, 2010, from http://en.wikipedia.org/wiki/IEEE

Neil H. Johnson *is an Assistant Professor in the Center for Language Research at the University of Aizu, Japan, where he teaches a variety of English for Specific Purposes courses. He earned his Ph.D. in Second Language Acquisition and Teaching from the University of Arizona.*

His main research interests are in sociocultural theory and discourse analysis. ***johnson@u-aizu.ac.jp***

William Rozycki *earned his Ph.D. at Indiana University in 1983. He is a professor at the University of Aizu and chair of the IEEE Professional Communication Society of Japan. His research focus is the presentation of engineering research findings by non-native speakers of English.* ***rozycki@u-aizu.ac.jp***

Manuscript received June 2, 2009
Revised version accepted April 2, 2010
Final version accepted September 2, 2010