Research paper

Authorial Identity Behind First Person Pronouns in English Research Articles of Native and Non-native Scholars

Cui Yu

College of Humanities and Law; China University of Mining and Technology, Beijing, China

Abstract

Few studies to date have investigated how native English scholars and non-native English scholars establish their authorial identity through first person pronouns. To this end, utilizing Işık-Taş’s (2018) discourse functions of first person pronouns as the analysis framework, it aims to examine how the authorial identity is represented by first person pronouns (I/me/my/we/us/our) in 40 English research articles of Applied Linguistics respectively written by native speaker scholars and non-native speaker scholars. Two sub-corpora were analyzed: native English speaker corpora and non-native English speaker corpora. The singular first person was found to be the preferred choice by both scholars. However, the analyses revealed differences in the distribution and discourse functions of first person pronouns. The first person occurred more frequently in native English speaker corpora than does non-native English speaker corpora. Based on Işık-Taş (2018) framework, low-risk functions (e.g., representing a community) and medium-risk functions (e.g., stating a goal) contrasted starkly between the two corpora. The variation in both corpora suggests that the use of first person pronouns in English research articles is not only associated with the cultural context but also by the author’s proficiency and competitiveness to publish paper internationally.

Keywords: authorial identity, first person pronoun, research article, discourse function

*Corresponding Author: Cui Yu

Address: College of Humanities and Law; China University of Mining and Technology, Beijing, China

Tel: (+86) 13366227625 E-mail: 17862347687@163.com
Introduction

When social interactions occur in the academic community, text is a place where knowledge and writer’s identity are constructed, negotiated, and created (Rahimivand & Kuhi, 2014). Writers use various linguistic resources, such as engagement (Chang & Schleppegrell, 2011), hedges (e.g., Yang, 2013), attitude markers (e.g., Hyland, 2005), and first person pronouns (e.g., Harwood, 2005; Işık-Taş, 2018) in research articles (RAs) to display their identity. Among these linguistic devices, first person pronouns are a central concern in the past few decades.

The motivation for the current study arises from previous studies (Hyland, 2002; Starfield & Ravelli, 2006; Matsuda, 2015). Some scholars work on authorial self-reference in English RAs by native speakers (NSE) and by speakers of other countries (e.g., Spanish, Turkey and Russia), but little research is concerned about Chinese. Moreover, they show how first-person pronoun is represented in RAs among different L1 writers’ background (Dueñas, 2007). A contrastive analysis on how EFL learners and native English speakers make use of first person pronouns in English RAs has been carried out (Işık-Taş, 2018). However, little research is conducted concerning authorial identity displayed among scholars.

This study responds to the need for additional research in this area, by examining the frequency and discourse functions of first person pronouns (I/me/we/us/my/our) in English RAs produced by NSE scholars, in comparison with a corpus of RAs produced by non-native speaker (NNSE) scholars, that is, Chinese scholars. The research questions are addressed in this study are:

1. What are the similarities and differences in the frequency of first person pronouns in English RAs written by NSE scholars and NNSE scholars in English-medium international journals?
2. What are the discourse functions behind the patterns of first personal use?

Literature Review

Identity in RAs

Identity is considered as a complex, contradictory and multifaceted notion (Norton, 1997), as well as some related words like “persona”, “self” and “voice” are used interchangeably by researchers in diverse disciplinary contexts (Rahimivand & Kuhi, 2014). It also involves phenomenological reality that exists in people’s perceptions (e.g., social constructs) (Matsuda, 2015). As a result, identity is hard to define. Specifically, identity in RAs, namely authorial identity, is the realization that writers construct a credible representation of themselves (Hyland, 2002).

According to Ivanič (1998), identity includes autobiographical self, discoursal self, self as author, and possibilities for self-hood in the socio-cultural and institutional context. To explain briefly, autobiographical self refers to “the identity of a person in the act of writing”. Discoursal self is described as “the impression which writers convey in a particular written text”. Self as author is explained as “the writer’s position, opinions and beliefs in writing”.

Cui. Authorial Identity Behind First Person Pronouns in English Research Articles of Native and Non-Native Scholars
Lastly, *possibilities for self-hood* is elaborated as “several socially available possibilities for self-hood and several ways of doing the same thing”.

Based on Ivanič (1998), Tang and John (1999) set up a typology of six different identities in academic writing: author as “representative of a group”, “the guide through the essay”, “the architect of the essay”, “the recounter of the research process”, “the opinion-holder” and “the originator of an idea”. Author as “representative of a group” shows the least powerful authorial presence, while author as “the originator of an idea” shows the most powerful one. Hyland (2002) posits four discourse functions for first person pronouns: stating a goal/purpose, explaining a procedure, elaborating an argument and stating results/claims.

Further, İşik-Taş (2018), integrating Tang and John (1999) taxonomy and Hyland (2002) framework, comes up with a new framework for discourse functions of first person pronouns, including three categories, namely low-risk functions (*representing a community and guiding the reader through the text*), medium-risk functions (*stating a goal/purpose and explaining a procedure*) and high-risk functions (*expressing an opinion, elaborating an argument, presenting a new idea/knowledge claim and stating results/claims*). Above definitions are given in the context of academic writing. While writing, writers have to select their words so that readers are drawn in, influenced and persuaded (Hyland, 2002). In this study, İşik-Taş (2018) is used as a starting point for the textual analyses of first-person pronouns due to its comprehensiveness and to present contrastive findings more clearly.

**First Person Pronouns in RAs**

First person pronouns are realized by means of self-references, i.e., singular and plural first person pronouns (*I/we/me/us*) and possessive adjectives (*my/our*) referring to the authors, and self-citations. As a powerful linguistic device, the interpretation of first person pronouns varies across the context and culture. They are found to be more common in English-language RAs published in international journals than in RAs in various other languages (Dueñas, 2007; Molino, 2010). Molino (2010), for example, took a contrastive approach and compared Linguistics RAs in English and Italian in terms of first-person subject pronouns. The results showed that personal forms were found to be less frequent in Italian Linguistics RAs. However, a few scholars come up with different opinions. For example, Sheldon (2009) explored the different identity behind first-person roles in English and Spanish RAs and found that the texts written in Spanish used slightly more personal forms than did the texts in English. The reason may lie in the globalization of economic exchange and scientific communication.

Moreover, first person pronouns have been investigated across various disciplines. Previous corpus-based studies found that different disciplines varied in use of pronouns, and that soft disciplines featured more first personal pronouns than the hard fields (Harwood, 2005; Hyland, 2001). For instance, Hyland (2001) focused on the use of self-citation and exclusive first person pronouns in a corpus of 240 RAs in 8 disciplines and concluded that an average of first person...
pronouns occurred in the humanities and social science papers, compared with 17 in science and engineering. Yet, to date, few studies have addressed the use of first person pronouns in the field of Applied Linguistics. In addition, relatively more studies have investigated first personal pronouns of English RAs by native and non-native learners’ writing (Çandarlı et al., 2015; Işık-Taş, 2018), but very few studies are concerned with first personal pronouns by scholars’ writing.

What’s more, of the first person pronouns, “I/we” pronouns were found to be more prevalently researched among scholars (Kuo, 1999; Martínez, 2005; Harwood, 2005, 2006; Dueñas, 2007). For example, Harwood (2006) described the use of the pronouns “I/we” in academic writing based on five political scientists’ interview-based accounts and the data revealed evidence of a lack of consciousness about certain aspects of pronoun use. However, other forms of first person pronouns, e.g., me/my/us/our, were less explored. As what has been seen above, what is needed is to further research to understand how NSE scholars and NNSE scholars use first personal pronouns (I/me/my/we/me/us) in their English RAs.

**Method**

**Corpus**

We began to categorize the scholars as NSE and NNSE by searching their names and institutions online, so the corpus for analysis consists of two sub-corpora, namely NSE corpus and NNSE corpus. The two sub-corpora represented RAs produced in English by native scholars and by Chinese scholars, each consisting of 40 single-authored RAs built with full text between 2015-2019 (See Table 1) (Appendix A). In addition, they were chosen from authoritative peer-reviewed journals. The NSE corpus was built with three highly-impacted international journals, including Language Learning, Applied Linguistics and Studies in Second Language Acquisition. The NNSE corpus was built with the only one published English journals in China, namely Chinese Journal of Applied Linguistics. In this study, the sub-corpora were representative of a genre--RA, a field--applied linguistics, and international standard.

The RAs were available online. They merely included main body while removing the title, abstract, acknowledgment, references, and appendix. In order to make the data more representative of authorial identity, direct quotations, interview, tables, figures and graphs were also removed. The count included only those tokens which expressed the scholars’ authorial voice, leaving aside engaging tokens (i.e., making joint reference to authors and scholars or generally to all disciplinary members) and self-references used in citations or in the reproduction of questionnaires (Dueñas, 2007).
Table 1
The size of the corpora

<table>
<thead>
<tr>
<th></th>
<th>Number of essays</th>
<th>Word Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>English essays by Chinese scholars</td>
<td>40</td>
<td>217,623</td>
</tr>
<tr>
<td>English essays by American scholars</td>
<td>40</td>
<td>329,386</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>549,009</td>
</tr>
</tbody>
</table>

Data Analysis

As stated previously, we took Işık-Taş (2018) framework as a starting point for analysis of discourse functions of first person pronouns (Table 2). This framework comprises three interrelated aspects of writer identity, including low-risk functions, medium-risk functions and high-risk functions. Low-risk functions are grouped into representing a community and guiding the reader through the text. The former refers to “a larger number of people”, while the latter means to draw the readers’ attention and arrive at a conclusion that the writer presumes is shared by the reader. Medium-risk functions are categorized into stating a goal/purpose and explaining a procedure. The former refers to signal intentions for texts, while the latter means to present methodological approach. High-risk functions involve expressing an opinion, elaborating an argument, presenting a new idea/knowledge claim and stating results/claim. Expressing an opinion refers to a personal statement, elaborating an argument refers to set out a line of reasoning, presenting a new idea/knowledge claim refers to originate new ideas, while stating results/claim refers to announce writers’ presence where they make a knowledge claim.

Table 2
Işık-Taş (2018) framework for analysis of discourse functions of first person pronouns

<table>
<thead>
<tr>
<th>Low-risk functions</th>
<th>Medium-risk functions</th>
<th>High-risk functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representing a community</td>
<td>Starting a goal/purpose</td>
<td>Expressing an opinion</td>
</tr>
<tr>
<td>Guiding the reader</td>
<td>Explaining a procedure</td>
<td>Elaborating an argument</td>
</tr>
<tr>
<td></td>
<td></td>
<td>through the text</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Presenting a new idea/knowledge claim</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Stating results/claim</td>
</tr>
</tbody>
</table>

We used the software Antcon3 3.4.3w to analyze the frequency of first-person pronouns--I, me, my, we, us and our--in the corpus by retrieving the concordance and UAM Corpus Tool 3 to code the discourse functions of first-person pronouns. To facilitate the comparison, we counted for each text per 1,000 words. Then the Log likelihood and Chi-square Calculator 1.0 was employed to analyze the significance test.

To realize the validity and reliability of the data, two lecturers majoring in Applied Linguistics who were familiar with the categories coded the discoursal functions, yielding an inter-rater reliability of 90%. To solve the cases of disagreement, we consulted a third rater,
who was also majoring in Applied Linguistics, to reach a consensus. After several discussions, full agreement was reached between the first and second categorization.

**Results and Discussion**

In our study, we found striking similarities and differences in the frequency and discourse functions of first person pronouns in RAs by NSE and NNSE scholars. The corpus-based findings showed that both scholars employed a higher number of 1st-person singular in their English RAs (Table 3). Additionally, based on Işık-Taş (2018) framework, low-risk functions and medium-risk functions of first person pronouns were notably different in English RAs by NSE and NNSE scholars (Table 4).

**Frequency of first person pronouns in the NSE and NNSE corpora**

As shown in Table 3 below, the overall frequency (per 1,000 words) of first person pronouns was 6.4. About 3.47 first person pronouns and 2.93 first person pronouns per 1,000 words were respectively employed in NSE and NNSE corpora. Clearly, the results were consistent with previous studies, that is, authorial reference was more frequently found in RAs in English than in RAs in other languages (Molino, 2010; Işık-Taş, 2018).

Among analyzed element, “I” was the most frequently appeared in both corpora, each with the frequency of 1.5 in NSE corpora and 1.22 in NNSE corpora. In addition, almost the identified first person pronouns in NSE corpora were 1st-person singular (2.04 per 1,000 words). Surprisingly, this was much similar with their counterparts in NNSE corpora, whereby the 1st-person singular (1.94 per 1,000 words) was found to be the preferred choice. Notably, the results contradicted previous studies, that is, the plural forms of the 1st-person pronoun were more commonly employed by NNSE writers because of modesty and collectivism (Chen, 2020). The possible explanation is that the study contrastively analyzes NSE corpora by native English scholars and NNSE corpora by Chinese scholars not by novice English learners. Because of globalization and academic competition, Chinese scholars struggle to approach the text genre of native speakers. And the second explanation might lie in the particularity of the text.

**Table 3**

*Use of first person pronouns per sub-corpus (per 1,000 words)*

<table>
<thead>
<tr>
<th>Sub-corpus</th>
<th>Total</th>
<th>I</th>
<th>Me</th>
<th>My</th>
<th>We</th>
<th>Us</th>
<th>Our</th>
</tr>
</thead>
<tbody>
<tr>
<td>NSE</td>
<td>3.47</td>
<td>1.5</td>
<td>0.24</td>
<td>0.3</td>
<td>0.77</td>
<td>0.22</td>
<td>0.44</td>
</tr>
<tr>
<td>NNSE</td>
<td>2.93</td>
<td>1.22</td>
<td>0.27</td>
<td>0.45</td>
<td>0.43</td>
<td>0.2</td>
<td>0.36</td>
</tr>
<tr>
<td>Total</td>
<td>6.4</td>
<td>2.72</td>
<td>0.51</td>
<td>0.75</td>
<td>1.2</td>
<td>0.42</td>
<td>0.8</td>
</tr>
</tbody>
</table>

**Discourse functions of first person pronouns**

According to Işık-Taş (2018), discourse functions include low-risk functions, medium-risk functions and high-risk functions. Low-risk functions consist of representing a community...
and *guiding the reader through the text*, because writers display an invisible authorial presence in their texts. Medium-risk functions are filled by *starting a goal/purpose* and *explaining a procedure* while high-risk functions are represented by *expressing an opinion*, *elaborating an argument*, *presenting a new idea/knowledge claim*, and *stating results/claim*. Table 4 showed that the percentage of first person pronouns in low-risk functions and medium-risk functions were significantly higher between the NSE and NNSE RAs. Low-risk functions and medium-risk functions by NSE scholars were almost the same (0.95 and 0.94 per 1000 words respectively). On the whole, pronouns that fulfilled high-risk functions were the least employed in the two corpora.

**Table 4**

Discourse functions of first person pronouns

<table>
<thead>
<tr>
<th>Functions</th>
<th>NSE sub-corpus</th>
<th>NNSE sub-corpus</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>representing a community</td>
<td>305 0.95</td>
<td>108 0.65</td>
<td>0.0000***</td>
</tr>
<tr>
<td>guiding the reader</td>
<td>263 0.82</td>
<td>100 0.6</td>
<td></td>
</tr>
<tr>
<td>Medium-risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>starting a goal/purpose</td>
<td>155 0.94</td>
<td>48 0.79</td>
<td>0.0000***</td>
</tr>
<tr>
<td>explaining a procedure</td>
<td>230 0.85</td>
<td>45 0.72</td>
<td></td>
</tr>
<tr>
<td>High-risk</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>expressing an opinion</td>
<td>176 0.63</td>
<td>37 0.33</td>
<td></td>
</tr>
<tr>
<td>elaborating an argument</td>
<td>34 0.12</td>
<td>5 0.04</td>
<td></td>
</tr>
<tr>
<td>presenting a new idea</td>
<td>52 0.19</td>
<td>12 0.11</td>
<td></td>
</tr>
<tr>
<td>stating results</td>
<td>35 0.13</td>
<td>4 0.04</td>
<td></td>
</tr>
<tr>
<td>***p&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Low-risk functions of first person pronouns**

In Tang and John’s framework (1999), inclusive first person pronouns (we/us/our) realize two authorial functions: “I” as the representative and “I” as a guide. The former usually refers to a large group of people, while the latter means that first person pronouns show the reader through text accompanied by verbs like see, observe and note. However, writers may become “a member of the audience”. Therefore, Işık-Taş (2018) further recasts the two functions as *representing a community* (Ex1) and *guiding the reader through the text* (Ex1). They are classified as low-risk functions, which are adopted in the study.

In Ex1, “we” might refer to a large number of people or part of the audience. In Ex2, the writer takes the reader through the text by explaining what they “observe”.

(Ex1): One thing we know is that the tasks administrated were both for building receptive and productive vocabulary knowledge. (CJAL7)
(Ex2): We observed how the participants oriented to their learning of the word from several different angles: from a pronunciation focus to a spelling focus. (AL11)

Due to the inclusive form of first person pronouns functioning as implicit authorial role, NSE scholars preferred to use in their texts. This result was line with Dueñas (2007) because of reducing writers’ personal intrusion and yet emphasizing the roles as authors (Hyland, 2001).

**Medium-risk functions of first person pronouns**

With the help of Hyland’s framework (2002), Işık-Taş (2018) proposes that *starting a goal/purpose* (Ex3) and *explaining a procedure* (Ex4) are categorized as medium-risk functions due to the fact that writers explicitly display authorial presence and are responsible for their research. The function of *starting a goal/purpose* refers to signal intentions and provided an overt structure for the texts. In Ex3, the writer shows the direction of the research and what he intends to do in the study.

(Ex 3): By modeling both systematicity and individuality simultaneously, I aim to demonstrate a more comprehensive view of morpheme accuracy development. (LL6)

*Explaining a procedure* refers to present methodological approach. As in Ex 4, the writer provides an account of research procedures by using ordinal number.

(Ex 4): In designing the spot dictation exercise, I first chose all the content words in the passage, then crossed out the same token words, finally selected the content words which covered as many phonemes as possible. (CJAL8)

Notably, the percentage of medium-risk functions in NSE RAs and NNSE RAs were found to contrast starkly. This was the most frequent use in the NNSE RAs while ranking the second in the NSE RAs. This may be because the ability to carry out research methodology is seen as a crucial element of RAs (Hyland, 2002) and NSE scholars recognize its importance.

**High-risk functions of first person pronouns**

According to Işık-Taş’s framework (2018), high-risk functions of first person pronouns include *expressing an opinion* (Ex 5), *elaborating an argument* (Ex 6), *presenting a new idea/knowledge claim* (Ex 7), and *stating results* (Ex 8).

In Tang and John’s taxonomy (1999), *Expressing an opinion* means that the person shares an opinion, view or attitude (e.g., by expressing agreement, disagreement or interest) with regard to known information or established facts, co-occurring with verbs depicting what Halliday (1994) terms mental processes of cognition. As in Ex5, the writer shares his opinion about the policy of de-emphasizing English in the Gaokao.
In *elaborating an argument*, the writers personally engage with their beliefs and their audience (Hyland, 2002). In Ex6, the writer shares his reasoning with the use of modal verbs “should”. Function of *presenting a new idea or knowledge claim* is represented by “I” as the originator (İşik-Taş, 2018). It signals writers’ new idea in the text as in Ex7.

**Stating results/claims**, as the most powerful role, explicitly foregrounds writers’ distinctive contribution and commitment to a position (Hyland, 2002). As in Ex8, the writer announces their findings about the brain.

(Ex 5): I personally think, if the policy of de-emphasizing English in the *Gaokao* were widely adopted, it would be time for Chinese stakeholders to de-emphasize the repetitious and tedious imitation of native English. (CJAL37)

(Ex 6): We should incorporate the principles of learners’ developmental readiness when selecting textbook content and sequencing the textbook’s syllabus. (CJAL38)

(Ex 7): We predict that the contributions of the dominance and proficiency factors to variability in Pillai scores will differ across language. (SSLA11)

(Ex 8): We found that the right brain had more electrodes activated than the left brain. (CJAL30)

Table 4 showed a similarity between the two corpora in the distribution of first person pronouns across high-risk functions. This role was the least preferred by both scholars. The reason may due to the fact that high-risk functions are a risky strategy that both scholars show preference to downplay authorial identity (Hyland, 2002). It should be pointed out, however, that NSE scholars seem to favour high-risk functions than do NNSE scholars. This result needs to be interpreted in the light of NSE scholars’ tendency to participate in the international academic discourse community (İşik-Taş, 2018).

Obviously, First person pronouns fulfilling low-risk and medium-risk functions by NSE scholars were almost the same (0.95 and 0.94 per 1000 words respectively), featuring the highest percentage. Therefore, it could be argued that NSE scholars are more likely to emphasize their role as researchers by means of low-risk and medium-risk functions. The percentage of first person pronouns in low-risk functions and medium-risk functions were significantly higher between the NSE RAs and NNSE RAs. The variability in how authorial identity is displayed by NSE and NNSE scholars in this study is likely attributable to different cultures and traditional view. However, in the English RAs produced by both NSE and NNSE scholars only a marginal portion of first person pronouns fulfilled high-risk functions (0.63 per 1000 words and 0.33 per 1000 words respectively). This may be because scholars will claim most powerful authorial presence by employing high-risk functions. Yet NSE scholars...
preferred high-risk functions than NNSE scholars due to their tendency to participate in the international academic discourse community. In light of these similarities and differences, we need to be aware of how academic convention position scholars to construct and reconcile their authorial identities. NNSE scholars could use more first person pronouns to front powerful and explicit authorial identity.

**Conclusion**

The study investigates how NSE scholars and NNSE scholars establish their authorial identities through first person pronouns. The results indicated that the NSE scholars presented their authorial role in their RAs more frequently than NNSE counterparts. The higher number of first person pronouns in the NSE RAs than NNSE one might be attributable to different conventional, traditional views on self-representation pertaining to the different national, big cultures (Atkinson, 2004). It should be pointed out that both scholars prefer to use singular first person pronouns in RAs. The reason could be probably explained as the high level of authors’ proficiency and competitiveness among scholars wishing to publish their RAs in an international journal (Dueñas, 2007).

Besides the greater distribution of use of first person pronouns in NSE corpora, the functions behind first person pronouns have yielded significant differences throughout RAs in NSE and NNSE. Low-risk functions and medium-risk functions are almost the same, becoming the most preferred choice by NSE scholars. In addition, low-risk functions and medium-risk functions that fulfilled authorial identity have been found to be significantly higher in the NSE corpora than in the NNSE one, while high-risk functions between the two corpora don’t contrast starkly. However, NSE scholars prefer to use more higher-risk functions than NNSE scholars. It suggests that the choice of NSE scholars tending to use higher-risk functions is influenced by their participation in the international academic discourse community (Işık-Taş, 2018).

Based on our results, we present evidence that the degree of authorial presence in RAs is determined not only by cultural context as has been demonstrated (Harwood, 2005; Dueñas, 2007; Chen, 2020), but also by authors’ proficiency and competitiveness to publish paper internationally, which is an interesting question. This is consistent with Işık-Taş (2018), suggesting that writers’ choice to publish internationally affects their linguistic choices in establishing authorial identity in writing.

This study has several limitations. First, because of the specialty and small size of the corpus, the results cannot be generalized to form a broader picture of first person pronouns use in writing. Moreover, what is worthy of noticing is that our focus is on a single discipline written by NSE scholars and NNSE scholars. As a result, researching authorial identity of other disciplines across various cultures is a must. Finally, this analysis of authorial identity behind first person pronouns is limited to corpus-based approach. More longitudinal studies and qualitative studies should be addressed by future studies.
References


**Appendix A. Research Articles in the corpora**

**A.1 NSE sub-corpora**


(LL5) Nakatsukasa, K. Efficacy of recasts and gestures on the acquisition of locative


NNSE sub-corpora


