

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

### Citation

Yu, C. (2021). Authorial identity behind first person pronouns in English research articles of native and non-native scholars. *Journal of new advances in English Language Teaching and Applied Linguistics*, 3(1), 462-478.

DOI: 10.22034/JELTAL.2021.129414

### Received

2021-01-21

### Revised

2021-02-18

### Accepted

2021-02-22

### Keywords:

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

### 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.

\*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".

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, Işık-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, Işık-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*

	Number of essays	Word Count
English essays by Chinese scholars	40	217,623
English essays by American scholars	40	329,386
Total	80	549,009

**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*

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

We used the software Antconc 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 discursal 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)*

Sub-corpus	Total	I	Me	My	We	Us	Our
NSE	3.47	1.5	0.24	0.3	0.77	0.22	0.44
NNSE	2.93	1.22	0.27	0.45	0.43	0.2	0.36
Total	6.4	2.72	0.51	0.75	1.2	0.42	0.8

#### 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*

Functions	NSE sub-corpus		NNSE sub-corpus		P
	N	/1000words	N	/1000words	
Low-risk	305	0.95	108	0.65	0.0000***
<i>representing a community</i>	263	0.82	100	0.6	
<i>guiding the reader</i>	42	0.13	8	0.5	
Medium-risk	155	0.94	48	0.79	0.0000***
<i>starting a goal/purpose</i>	25	0.09	3	0.05	
<i>explaining a procedure</i>	230	0.85	45	0.72	
High-risk	176	0.63	37	0.33	
<i>expressing an opinion</i>	34	0.12	5	0.04	
<i>elaborating an argument</i>	52	0.19	12	0.11	
<i>presenting a new idea</i>	35	0.13	4	0.04	
<i>stating results</i>	55	0.2	6	0.14	

\*\*\*p<0.001

### 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 (Işık-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 (Işık-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

- Atkinson, D. (2004). Contrasting rhetorics/contrasting cultures: Why contrastive rhetoric needs a better conceptualization of culture. *Journal of English for Academic Purposes*, 3(4), 277–289.
- Çandarlı, D., Bayyurt, Y., & Martı, L. (2015). Authorial presence in L1 and L2 novice academic writing: Cross-linguistic and cross-cultural perspectives. *Journal of English for Academic Purposes*, 20, 192-202.
- Chang, P., & Schleppegrell, M. (2011). Taking an effective authorial stance in academic writing: Making the linguistic resources explicit for L2 writers in the social sciences. *Journal of English for Academic Purposes*, 10(3), 140-151.
- Chen, R. (2020). Single author self-reference: Identity construction and pragmatic competence. *Journal of English for Academic Purposes*, 45, 1-14.
- Dueñas, P. (2007). ‘I/we focus on...’: A cross-cultural analysis of self-mentions in business management research articles. *Journal of English for Academic Purposes*, 6(2), 143-162.
- Harwood, N. (2005). ‘I hoped to counteract the memory problem, but I made no impact whatsoever’: Discussing methods in computing science using ‘I’. *English for Specific Purposes*, 24(3), 243–267.
- Harwood, N. (2006). (In)appropriate personal pronoun use in political science: A qualitative study and a proposed heuristic for future research. *Written Communication*, 23(4), 424-445.
- Hyland, K. (2001). Humble servants of the discipline: Self-mention in research articles. *English for Specific Purposes*, 20, 207–226.
- Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics*, 34(8), 1091-1112.
- Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, 7(2), 173-192.
- İşik-Taş, E. (2018). Authorial identity in Turkish language and English language research articles in Sociology: The role of publication context in academic writers' discourse choices. *English for Specific Purposes*, 49, 26–38.
- Ivanič, R. (1998). *Writing and identity: The discursive construction of identity in academic writing*. Amsterdam: John Benjamins
- Kuo, C.-H. (1999). The use of personal pronouns: Role relationships in scientific journal articles. *English for Specific Purposes*, 18(2), 121-138.
- Martínez, I. A. (2005). Native and non-native writer’s use of first person pronouns in the different sections of biology research articles in biology. *Journal of Second Language Writing*, 14(3), 174–190.
- Matsuda, P. K. (2015). Identity in written discourse. *Annual Review of Applied Linguistics*, 35, 140–159.
- Molino, A. (2010). Personal and impersonal authorial references: A contrastive study of English and Italian linguistics research articles. *Journal of English for Academic Purposes*, 9, 86-101.

- Norton, B. (1997). Language, identity, and the ownership of English. *Tesol Quarterly*, 31(3), 409-429.
- Rahimivand, M., & Kuhi, D. (2014). An Exploration of Discoursal Construction of Identity in Academic Writing. *Procedia-Social and Behavioral Sciences*, 98, 1492–1501.
- Sheldon, E. (2009). From one I to another: Discursive construction of self-representation in English and Castilian Spanish research articles. *English for Specific Purposes*, 28(4), 251-265.
- Starfield, S., & Ravelli, J. (2006). “The writing of this thesis was a process that I could not explore with the positivistic detachment of the classical sociologist”: Self and structure in New Humanities research theses. *English for Specific Purposes*, 5(3), 222–243.
- Tang, R., & John, S. (1999). The ‘I’ in identity: exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes*, 18(1), 23-39.
- Yang, Y. (2013). Exploring linguistic and cultural variations in the use of hedges in English and Chinese scientific discourse. *Journal of Pragmatics*, 50(1), 23-36.

## Appendix A. Research Articles in the corpora

### A. 1. NSE sub-corpora

- (AL1) Deroey, K. L. (2015). Marking importance in lectures: Interactive and textual orientation. *Applied linguistics*, 36(1), 51-72.
- (AL2) Barbieri, F. (2015). Involvement in university classroom discourse: Register variation and interactivity. *Applied linguistics*, 36(2), 151-173.
- (AL3) Riazi, A. M. (2016). Innovative mixed-methods research: Moving beyond design technicalities to epistemological and methodological realizations. *Applied Linguistics*, 37(1), 33-49.
- (AL4) Trebits, A. (2016). Sources of individual differences in L2 narrative production: The contribution of input, processing, and output anxiety. *Applied linguistics*, 37(2), 155-174.
- (AL5) Alrabai, F. (2016). The effects of teachers’ in-class motivational intervention on learners’ EFL achievement. *Applied linguistics*, 37(3), 307-333.
- (AL6) Durrant, P. (2017). Lexical bundles and disciplinary variation in university students’ writing: Mapping the territories. *Applied Linguistics*, 38(2), 165-193.
- (AL7) Negretti, R. (2017). Calibrating genre: Metacognitive judgments and rhetorical effectiveness in academic writing by L2 graduate students. *Applied Linguistics*, 38(4), 512-539.
- (AL8) Hawkins, M. R. (2018). Transmodalities and transnational encounters: Fostering critical cosmopolitan relations. *Applied Linguistics*, 39(1), 55-77.
- (AL9) Nguyen, H. T. (2018). Interactional practices across settings: From classroom role-plays to workplace patient consultations. *Applied Linguistics*, 39(2), 213-235.
- (AL10) Nagle, C. L. (2019). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*, 40(1), 86-107.
- (AL11) Anderson, T. (2019). Reproductions of Chinese transnationalism: Ambivalent identities in study abroad. *Applied Linguistics*, 40(2), 228-247.

- (AL12) Sifakis, N. C. (2019). ELF awareness in English language teaching: Principles and processes. *Applied Linguistics*, 40(2), 288-306.
- (AL13) Toohey, K. (2019). The onto-epistemologies of new materialism: implications for applied linguistics pedagogies and research. *Applied Linguistics*, 40(6), 937-956.
- (LL1) Casasanto, D. (2016). A shared mechanism of linguistic, cultural, and bodily relativity. *Language Learning*, 66(3), 714-730.
- (LL2) Donaldson, B. (2017). Negation in near-native French: Variation and sociolinguistic competence. *Language Learning*, 67(1), 141-170.
- (LL3) Jegerski, J. (2018). Sentence processing in Spanish as a heritage language: A self-paced reading study of relative clause attachment. *Language Learning*, 68(3), 598-634.
- (LL4) Pienemann, Manfred (2015). An outline of processability theory and its relationship to other approaches to SLA. *Language Learning*, 65(1), 123–151.
- (LL5) Saito, K. (2015). Experience effects on the development of late second language learners' oral proficiency. *Language Learning*, 65(3), 563-595.
- (LL6) Thomas, Margaret (2015). Air writing as a technique for the acquisition of Sino-Japanese characters by second language learners. *Language Learning*, 65(3), 631–659.
- (LL7) Ryan, J. (2015). Overexplicit referent tracking in L2 English: Strategy, avoidance, or myth. *Language Learning*, 65(4), 824-859.
- (LL8) Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, 66(4), 834-871.
- (LL9) Vanhove, J. (2017). The influence of standard and substandard Dutch on gender assignment in second language German. *Language Learning*, 67(2), 431-460.
- (LL10) Shimanskaya, E. (2018). On the role of input in second language acquisition: The case of French strong pronouns. *Language Learning*, 68(3), 780-812.
- (LL11) Hulstijn, J. H. (2019). An individual-differences framework for comparing nonnative with native speakers: Perspectives from BLC theory. *Language Learning*, 69(1), 157-183.
- (LL12) Blom, E. (2019). Domain-general cognitive ability predicts bilingual children's receptive vocabulary in the majority language. *Language Learning*, 69(2), 292-322.
- (LL13) Uchikoshi, Y. (2019). Phonological awareness trajectories: Young Spanish–English and Cantonese–English bilinguals. *Language Learning*, 69(4), 802-838.
- (LL1) Stringer, D. (2015). Embedded wh-questions in L2 English in India: Inversion as a main clause phenomenon. *Studies in Second Language Acquisition*, 37(1), 101-133.
- (LL2) Miller, A. K. (2015). Intermediate traces and intermediate learners: Evidence for the use of intermediate structure during sentence processing in second language French. *Studies in Second Language Acquisition*, 37(3), 487-516
- (LL3) Bowden, H. W. (2016). Assessing second-language oral proficiency for research. *Studies in Second Language Acquisition*, 38(4), 647-675.
- (LL4) Shport, I. A. Training English listeners to identify pitch-accent patterns in Tokyo Japanese. *Studies in Second Language Acquisition*, 2016, 38(4): 739–769.
- (LL5) Nakatsukasa, K. Efficacy of recasts and gestures on the acquisition of locative

- prepositions. *Studies in Second Language Acquisition*, 2016, 38(4), 771-799.
- (LL6) McKinnon, S. (2017). TBLT instructional effects on tonal alignment and pitch range in L2 Spanish imperatives versus declaratives. *Studies in Second Language Acquisition*, 39(2), 287-317.
- (LL7) Solon, M. (2017). Do learners lighten up. *Studies in Second Language Acquisition*, 39(4), 301-332.
- (LL8) Foote, R. (2017). The storage and processing of morphologically complex words in L2 Spanish. *Studies in Second Language Acquisition*, 39(4), 735-767.
- (LL9) Mohamed, A. A. (2018). Exposure frequency in L2 reading: An eye-movement perspective of incidental vocabulary learning. *Studies in Second Language Acquisition*, 40(2), 269-293.
- (LL10) Malone, J. (2018). Incidental vocabulary learning in SLA. *Studies in Second Language Acquisition*, 40(3), 651-675.
- (LL11) Papi, M. (2018). Motivation as quality: regulatory fit effects on incidental vocabulary learning. *Studies in Second Language Acquisition*, 40(4), 707-730.
- (LL12) Shea, C. (2019). Dominance, proficiency, and Spanish heritage speakers' production of English and Spanish vowels. *Studies in Second Language Acquisition*, 41(1), 1-27.
- (LL13) Granena, G. (2019). Cognitive aptitudes and L2 speaking proficiency: Links between LLAMA and Hi-LAB. *Studies in Second Language Acquisition*, 41(2), 313-336.
- (LL14) Barkaoui, K. (2019). What can L2 writers' pausing behavior tell US about their L2 writing process. *Studies in Second Language Acquisition*, 41(3), 529-554.

### **NNSE sub-corpora**

- (CJAL1) Chen Sibó. (2015). The development of bilingual children's narrative skills: A report of the "Looking Glass Neighborhood" Program. *Chinese Journal of Applied Linguistics*, 38(3), 301-318.
- (CJAL2) Hu Jian. (2015). Interaction in assessment-oriented role-play: A conversation analytic approach. *Chinese Journal of Applied Linguistics*, 38(4), 472-489.
- (CJAL3) Li Di. (2015). Comparison between peer mediation and teacher mediation in dynamic EFL writing assessment: A case study. *Chinese Journal of Applied Linguistics*, 38(4), 490-502.
- (CJAL4) Li Jin. (2015). Developing Chinese complimenting in a study abroad program. *Chinese Journal of Applied Linguistics*, 38(3), 277-300.
- (CJAL5) Li Shunying. (2015). A case study of the English teachers' burnout in a Medical University in China. *Chinese Journal of Applied Linguistics*, 38(2), 234-245.
- (CJAL6) Lin Meizhen. (2015). Metadiscourse use in English job applications. *Chinese Journal of Applied Linguistics*, 38(3), 355-367.
- (CJAL7) Wang Ruixin. (2015). Task-induced involvement load: The accumulation effects on vocabulary acquisition. *Chinese Journal of Applied Linguistics*, 38(2), 150-165.
- (CJAL8) Zhang Lingli. (2015). An empirical study on the intelligibility of English spoken by Chinese university students. *Chinese Journal of Applied Linguistics*, 38(1), 36-54.

- (CJAL9) Zhang Qianmei. (2015). A study on Chinese students' perceptual learning styles, ideal L2 self, and L2 motivated L2 behavior. *Chinese Journal of Applied Linguistics*, 38(1), 93-109.
- (CJAL10) Chu Yan. (2016). A Contrastive Pragmatic Study of Politeness Strategies in Disagreement between Native Speakers of English and Chinese EFL Learners. *Chinese Journal of Applied Linguistics*, 39(2), 231-245.
- (CJAL11) Ding Xiaowei. (2016). Problem-Based L2 Learning: Self-Negotiated Linguistic Cognition. *Chinese Journal of Applied Linguistics*, 39(3), 316-336.
- (CJAL12) Ding Yu. (2016). Patterns of peer review in an EFL writing course in China. *Chinese Journal of Applied Linguistics*, 39(1), 72-91.
- (CJAL13) Li Qinghua. (2016). Chinese EFL Learners' Perceptions of Their Experience with Portfolio-Based Writing Assessment: A Case Study. *Chinese Journal of Applied Linguistics*, 39(2), 215-230.
- (CJAL14) Song Tiehua. (2016). Study on Chinese college EFL learners' comprehension and production of English relative clauses: testing three hypotheses. *Chinese Journal of Applied Linguistics*, 39(1), 26-37.
- (CJAL15) Teng Feng. (2016). Development of morphological awareness through English songs: A case study. *Chinese Journal of Applied Linguistics*, 39(2), 167-184
- (CJAL16) Wang Chaochang. (2016). How teachers' beliefs are in accordance with major SLA notions. *Chinese Journal of Applied Linguistics*, 39(3), 354-373.
- (CJAL17) Wang Qian. (2016). Embodied cognition processing and representation of power words by second language learners with different proficiency levels. *Chinese Journal of Applied Linguistics*, 39(4), 484-494.
- (CJAL18) Wu Shinian. (2016). Social and political dimensions of communication in grammar instruction: What teachers should contemplate. *Chinese Journal of Applied Linguistics*, 39(2), 153-16.
- (CJAL19) Xu Chengping. (2016). Acquisition of English reflexive verb constructions: Roles of syntactical indeterminacy and token frequency. *Chinese Journal of Applied Linguistics*, 39(3), 251-268.
- (CJAL20) Yuan Surong. (2016). The functional approach to the teaching of translation: A case study on Beiyong and its English translations. *Chinese Journal of Applied Linguistics*, 39(3), 304-315.
- (CJAL21) Zhang Jianying. (2016). A comparison of the L2 motivational self system Between Chinese EFL and ESL Learners. *Chinese Journal of Applied Linguistics*, 39(3), 269-287.
- (CJAL22) Chen Jianlin. (2017). Factors affecting Chinese raters' rating of high-stakes English exam essays. *Chinese Journal of Applied Linguistics*, 40(2), 208-228.
- (CJAL23) Chen Weirong. (2017). Functions of *So* in English as a lingua franca and native speech: A corpus-Based study. *Chinese Journal of Applied Linguistics*, 40(1), 94-111.
- (CJAL24) Dong Yanning. (2017). Teaching and assessing critical thinking in second language writing: An infusion approach. *Chinese Journal of Applied Linguistics*, 40(4), 431-451.
- (CJAL25) Li Danli. (2017). Autonomy in scaffolding as learning in teacher-student negotiation of meaning in a university EFL classroom. *Chinese Journal of Applied Linguistics*, 40(4), 410-

430.

(CJAL26) Lin Lin. (2017). Cultural flows and pedagogical dilemmas: Teaching with collaborative learning in the Chinese HE EFL context. *Chinese Journal of Applied Linguistics*, 40(1), 21-41.

(CJAL27) Wang Yi. (2017). When teacher autonomy meets management autonomy to enhance learner autonomy. *Chinese Journal of Applied Linguistics*, 40(4), 392-409.

(CJAL28) Lin Meixiao. (2018). Oral English skills of Chinese university students in describing their native culture. *Chinese Journal of Applied Linguistics*, 41(3), 338-355.

(CJAL29) Luo Han. (2018). Predictors of foreign language anxiety: A study of college-level L2 learners of Chinese. *Chinese Journal of Applied Linguistics*, 41(1), 3-24.

(CJAL30) Wang Qian. (2018). Neural mechanism and representation of English and Chinese metaphors of bilinguals with different second language proficiency: An ERP study. *Chinese Journal of Applied Linguistics*, 41(1), 67-83.

(CJAL31) Bao Wen. (2019). On teachers' professional qualities structure and development mechanisms in the business English discipline. *Chinese Journal of Applied Linguistics*, 42(2), 218-235.

(CJAL32) Chen Yaping. (2019). Developing students' critical thinking and discourse level writing skill through teachers' questions: A sociocultural approach. *Chinese Journal of Applied Linguistics*, 42(2), 141-162.

(CJAL33) Huang Enmou. (2019). Teaching to, with, and against the test: Language teaching and teacher identity under institutional neoliberalization. *Chinese Journal of Applied Linguistics*, 42(3), 307-326.

(CJAL34) Huang Zheng. (2019). An exploratory study of non-native English-speaking teachers' professional identity construction in a globalizing China. *Chinese Journal of Applied Linguistics*, 42(1), 40-59.

(CJAL35) Lin Jia. (2019). Factors related to EFL/ESL learners' reading strategy use: A literature review. *Chinese Journal of Applied Linguistics*, 42(1), 92-112.

(CJAL36) Ruan Zhoulin. (2019). Metadiscourse use in L2 student essay writing: A longitudinal cross-contextual comparison. *Chinese Journal of Applied Linguistics*, 2019, 42(4): 466-487.

(CJAL37) Si Jinghui. English as a lingua franca: A new approach for English language teaching in China. *Chinese Journal of Applied Linguistics*, 42(1), 113-138.

(CJAL38) Tang Xiaofei. (2019). Learnability of grammatical sequencing: A processability perspective of textbook evaluation in EFL settings. *Chinese Journal of Applied Linguistics*, 42(2), 236-257.

(CJAL39) Xu Ying. (2019). Scaffolding students' self-assessment of their English essays with annotated samples: A Mixed-Methods Study. *Chinese Journal of Applied Linguistics*, 42(4), 503-524.

(CJAL40) Wang Youkun. (2019). The relationship among non-English majors' autonomous learning ability, learning burnout and English grades: A structural equation modeling study. *Chinese Journal of Applied Linguistics*, 42(1), 79-91.