



A Comparative Study of Text Formality of Applied Linguistics Articles Written in English by Iranian and Native Speaking Researchers

Research Article
pp. 7-25

Sasan Baleghizadeh*¹
Faria Asadi²

Received: 08/04/2021

Accepted: 26/10/2021

Abstract

International scientific communication is mostly conducted in English. Consequently, writing and publishing in English is of significant importance in academic settings. Alongside many other factors, proper academic writing has an appropriate level of linguistic formality. Research has shown that linguistic features can distinguish between formal and informal texts. Different scholars have utilized different methods to define and measure formality. The present study compares the degree of formality of applied linguistics articles written in English by native English speaking and Iranian non-native English-speaking researchers by calculating their F-scores, a measure of formality introduced by Heylighen and Dewaele (1999). In total, 80 articles were selected from four international journals. Half of them were written by Iranian non-native researchers and the other half by native researchers. The results indicated a medium level of formality in both groups. However, articles written by Iranian non-native researchers were found to have a significantly higher degree of formality. Broadly, this research has implications for teachers of English in different areas, journal editors, materials developers and researchers who want to publish internationally.

Keywords: formality, native English-speaking researchers, non-native English-speaking researchers, applied linguistics, research article

* Corresponding Author

¹ Associate Professor, Department of English Language and Literature, Faculty of Letters & Human Sciences, Shahid Beheshti University, Tehran, Iran. s_baleghizadeh@sbu.ac.ir

² MA in TEFL, Department of English Language and Literature, Faculty of Letters & Human Sciences, Shahid Beheshti University, Tehran, Iran. faria.asadi@mail.sbu.ac.ir

DOI: 10.22051/lghor.2021.30138.1472

DOR: 20.1001.1.2588350.2022.6.3.1.6

Introduction

There is no doubt that publication is the dominant way in which scholars can share the findings of their research with other members of their scholarly community. Publishing in international journals which have a wide international audience can expand the readership of papers. More specifically, publishing in English, which is the main language of scientific research, results in wider readership and recognition of articles. In order to be successful in publishing in prestigious international journals written in the medium of English, in addition to a high language proficiency, researchers need to have a good command of the rules and trends that govern academic writing in English. One of the many characteristics that a proper academic paper possesses is an appropriate level of language formality. The need for academic writing to be formal is not a new subject and has been discussed for a long time. In practice, however, following the rules of formality is not that straightforward. What further complicates the matter is the situational flexibilities that linguistic formality allows and sometimes even values. Not knowing how to benefit from rigidity and flexibility of formal language in a balanced manner may result in rejection of papers by well-known journals. Nowadays, with the undeniable increased use of the Internet, the line between formal and informal language may be getting blurry since many of the contents on the Internet (e.g., blog posts) are usually informal, while academic writing (e.g., research article) is formal. As stated by Pavlick and Tetreault (2016), “the ability to recognize and respond to differences in formality is a necessary part of full language understanding” (p. 72).

Language, as Mesthrie et al. (2009) state, is “said to be indexical of one’s social class, status, region of origin, gender, age group and so on” (p. 6). Consequently, it can be said that different language backgrounds lead to different language uses, including linguistic formality.

In the previous studies in this area, different methods were introduced and operationalized in measuring/assessing the formality level of different types of texts.

The present study compared the applied linguistics articles written by native and Iranian non-native researchers which are published in international journals whereas in similar studies, the Iranian samples were chosen from Iranian local journals. Therefore, this study was conducted in order to find out whether the level of formality adopted by native and non-native researchers vary on a wider international scale.

Review of the Related Literature

Background

Nowadays, there is an increasing focus on professional academic writing around the world alongside a growing pressure on researchers to publish in the medium of English (Canagarajah, 1996, 2002; Curry & Lillis, 2004; Flowerdew, 1999; Hryniuk, 2015; Lillis & Curry, 2006; Mur-Duenas & Jolanta, 2016; Tardy, 2004). Such a condition is making it hard to avoid concluding that words “international” and “English” are becoming synonymous in academic writing (Lillis & Curry, 2010). Although researchers have not stopped publishing in their national languages (Belcher & Connor, 2001;

Canagarajah, 2002; Curry & Lillis, 2004), the benefits of it pushes scholars to want to publish in English international publications.

In general, it has been proved in previous research that linguistic features such as choice of words can differentiate between formal and informal language (Biber, 1988, 1995; Biber & Conrad, 2019; Conrad & Biber, 2001; Graesser et al., 2014; Graesser et al., 2011; Sardinha & Pinto, 2014). This has been confirmed in different languages, such as English (e.g., Biber, 1988), Korean (Kim & Biber, 1994), Somali (Biber & Hared, 1992), and Spanish (Biber et al., 2006), and in specialized areas, such as academic writing, university textbooks, and university student papers (Biber & Conrad, 2019).

Lillis (2013) maintains that “the binary between what constitutes formal/informal seems to be in widespread use in sociolinguistic texts yet is rarely precisely defined” (p. 9). Hyland and Jiang (2017) highlight the same point by stating that informality is “a slippery concept, difficult to pin down with a clear definition. It is typically either defined in contrast with formality, or in terms of lists of language features which are thought to comprise informal elements” (pp. 40-41).

Abu Sheikha and Inkpen (2010) claim that “vocabulary choice is perhaps the biggest style marker” (Introduction section). Liardét et al. (2019) found that usage of appropriate vocabulary is among the two most significant characteristics of formal language. In a nutshell, when confronted with a formal situation, people shift from informal words to equivalent formal ones. This “vocabulary shift is indeed the most salient feature of academic writing” as stated by Swales and Feak (2012, p. 14).

Abu Sheikha and Inkpen (2010) found features that influence linguistic style of a text and are good indicators of formal and informal language. Table 1 lists the features they found ordered from the strongest to the weakest.

Table 1

Formality Attributes and Their Weights by Abu Sheikha and Inkpen (2010)

Attributes	Their Weight
Informal pronouns	0.9031
Word length's average;	0.7729
Formality: (complex words, high value)	
Informality: (simple words, small value)	
Informal words list	0.4153
Active voice (informal)	0.3159
Contractions (informal)	0.2697
Type Tokens Ratio (TTR); TTR is the number of distinct words in a text comparing to the total number of words. In formal texts its value is lower than in informal texts.	0.1523
Passive Voice (formal)	0.1174
Abbreviations (informal)	0.0967
Phrasal Verbs (informal)	0.0735
Formal words list	0.057
Formal Pronouns	0.0183

Formality, as stated by Heylighen and Dewaele (1999), is one dimension of style which is seen as the most important aspect of variation in language. Heylighen and Dewaele (1999) introduced contextuality as

fundamental to formality. They formulated a quantification method which is based on context-dependence of the word categories in a text. This formula is known as the F-score and measures the formality of the whole text. Based on this context-dependence concept, an expression is formal when it has a “stable, recognizable form” and an invariant meaning that does not change when it is used “at different times, in different situations, or by different people, that is to say in different contexts” (Heylighen, 1999, p. 27). As Heylighen (1999) further argues, context-dependence, which is the opposite of formality, states that it is the context of the expression (e.g., speaker, audience, situation, etc.) that determines its meaning.

Pavlick and Tetreault (2016) mention that the F-score (Heylighen & Dewaele, 1999) is perhaps the best-known measure of formality for larger units of language. Moreover, as Nowson et al. (2005) assert, “it is Heylighen and Dewaele’s F-measure which has been used specifically to investigate individual differences between writers within a genre” (p. 1667). Therefore, this measure is utilized in the present study.

As Heylighen and Dewaele (1999) argue, where an expression lies on the continuum of formality is determined by choices made by the speaker/writer. As Heylighen and Dewaele (2002) argue, the variation of context-dependence found between different styles is “apparently due to the personal preferences of the subjects” and explains “more than half of the variance” between them (p. 14).

F-Score Formula

Conducting a statistical factor analysis on different word classes in French, Dewaele (1996) found that the number of nouns, adjectives, articles, and prepositions varied positively with formal requirements of situation and negatively with the number of verbs, pronouns, adverbs and interjections. Heylighen and Dewaele (1999) claim that nouns, adjectives, articles and prepositions are usually used to create context-independent expressions, while verbs, pronouns, adverbs and interjections usually build context-dependent expressions. Based on this proposition, Heylighen and Dewaele (1999) introduced the following formula as a measure of formality:

$$F = \frac{\text{noun freq.} + \text{adjective freq.} + \text{preposition freq.} + \text{article freq.} - \text{pronoun freq.} - \text{verb freq.} - \text{adverb freq.} - \text{interjection freq.} + 100}{2}$$

Applying the formula to data about frequency of word classes in different languages confirmed this hypothesis. Heylighen and Dewaele (1999) did factor analyses on existing data on seven different languages and found that a factor similar to the F-score appears as the most distinguishing factor in them. In other words, factor analyses on seven languages, including English, showed that there is a specific factor which explains the majority of the variation observed between different samples. Nouns, determiners and prepositions obtained positive loadings on that factor, whereas pronouns, adverbs, and verbs obtained negative loadings.

Peterson et al. (2011) suggested that comparing the behavior between different social groups can give some information about their culture by exploring “communication techniques such as formality” and identifying the norms and outliers of behavior within that social group (p. 94). Chambers et al.

(2003) assert that different social, educational and cultural backgrounds lead to different writing styles. Similarly, as mentioned earlier, language is “said to be indexical of one’s social class, status, region of origin, gender, age group and so on” (Mesthrie et al., 2009, p. 6). Drawing on such propositions which indicate that language can divide people into different groups, it can be concluded that native and non-native speakers of English form two distinct groups.

Although it has been argued that native-speaker norms should no longer be the standard form (e.g., Jenkins, 2003, 2006), what is actually observed is that native-speaker norms are still preferable in some cases (e.g., Celik, 2006; Florence Ma, 2012; Jenkins, 2011; Ozturk & Atay, 2010; Strauss, 2017, 2019). For example, Strauss (2017) herself condemns forcing non-native speakers to follow native-speakers’ accepted norms. However, she also confirms that “a number of highly respected researchers believe that the English currently used in prestigious journals is the best vehicle in which to report research. They are concerned that changes to the language will impact on the clarity and succinctness of such writing” (Strauss, 2017, p. 6). In Strauss’s (2019) study, in the same vein, international journal reviewers were interviewed and some of them recommended that authors consult native speakers. Hence, as Moreno (2010) states, in such conditions “protesting against and criticizing mainstream practices would be a disservice to the scholars” (p. 58). Therefore, it can be concluded that non-native speakers of English need to be aware of the appropriate and accepted English in international publishing.

The Present Study

The present study sought to answer the following research questions:

RQs: What is the degree of formality of applied linguistics articles written in English by native and Iranian non-native writers? Are they significantly different? If yes, which group (native or non-native) writes more formally?

Methodology

Corpus

Eighty articles were randomly selected from RELC Journal, SYSTEM, Language Teaching Research, and Language Learning Journal. These journals are internationally well-known in the field of applied linguistics and have almost the same prestigious level. They are among journals in which Iranian scholars have published their papers.

There were 4 journals in total, and 20 articles were selected from each: 10 written by native writers and 10 by Iranian non-native researchers. That is to say, 40 of them were written by Iranian non-native speakers of English and 40 by native speakers of English in total. The reason for selecting all 20 articles written by both native and non-native scholars from the same journal was that some differences are due to journal guidelines for publishing not because of some differences among the authors. Therefore, by minimizing the number of journals and having a more homogenous sample, the risk of having a false difference was minimized as well.

The articles were selected randomly according to the criteria mentioned below:

- They were published between 2010 and 2020.

- They were written solely by non-native authors, not co-written by native ones or vice versa.
- They had one to three authors.
- They were accessible online.
- Different articles that were written by the same author were excluded because some authors have their own style of writing that may affect the sample and bias the results. In other words, articles that were written (or co-written) by same authors were not included.
- The authors' affiliations were checked. Regarding the Iranian non-native sample, only articles which were submitted from Iran were included. The same procedure was followed for native authors and articles submitted from English-speaking countries were selected.

Procedure

As mentioned earlier, all the 80 articles were retrieved online and downloaded in PDF format. In order to compute the formality of the 80 articles, some parts of them needed to be discarded including references, acknowledgments, footnotes, tables, direct block quotations, sample questions of questionnaires, interview excerpts, appendices, declaration of conflicting interests, and funding.

In order to do so, all the PDF files were converted into Microsoft Word format and the unwanted parts were deleted manually. In case there were some spelling errors caused after the conversion, all word files were read and edited manually using Review feature of Word.

Then, the texts of these Word files were copied into TagAnt_64bit (version 1.2.0), a freeware (non-commercial) Part-Of-Speech (POS) tagger¹. This program is built on TreeTagger, which is a tool for annotating text with part-of-speech and lemma information developed by Helmut Schmid. It has been successfully used to tag many languages including English. Here, TagAnt analyzed the texts and labeled the words according to their part of speech. After having the words of the texts tagged, the word classes needed for computing the F-score introduced by Heylighen and Dewaele (1999) were selected. These word classes are: nouns, adjectives, prepositions, articles, pronouns, verbs, adverbs and interjections. The F-score formula can be found below:

$$F = \frac{\text{noun freq.} + \text{adjective freq.} + \text{preposition freq.} + \text{article freq.} - \text{pronoun freq.} - \text{verb freq.} - \text{adverb freq.} - \text{interjection freq.} + 100}{2}$$

As the next step, in every article, the number of words belonging to each part of speech was counted by a Python software programmed by Farzad Asgari². The resulting information was inserted into the F-score formula using Microsoft Excel. The frequency of each part of speech in this formula is expressed as percentage with respect to the total number of words in each text. The same procedure was followed for all the 80 articles and the resulting F-scores were compared. Heylighen and Dewaele (1999) asserted that this formula can be used for comparison within the same language and on samples

¹ Anthony, L. (2015). TagAnt (Version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from <https://www.laurenceanthony.net/software>

² Farzad Asgari is a Master student of Hydraulic Engineering at Kharazmi University of Tehran, Iran. std_farzad.asgari@khu.ac.ir

containing at least a few hundred words. These two criteria were met in the present study.

TagAnt does not tag articles (a, an, the) separately, instead, it includes all of them in a broader category alongside other types of determiners. To find the frequency of each of these articles (a, an, the), the text of all the 80 articles were analyzed using the same Python program. The number of each definite and indefinite article was counted and then added up for each research article.

Results

Human beings, as Bei (2011) mentioned, “appear to boast the ability to distinguish intuitively between different degrees of formality in language” (p. 32). As a student of English language who has continuously been exposed to English texts for about 6 years, the present researcher intuitively hypothesized that in general, English articles written by non-native Iranian authors seem to be more formal. Intuition, however, appears less helpful when more accurate explanation is required. Therefore, by using the F-score as a proven measure of formality, this was checked in this study. The arguments for choosing this measure is elaborated below.

Bei (2011) found that the F-score is a sensitive measure when it comes to detecting subtle differentiations in different situations and “a great extent of precision could be reached in measuring formality, especially with the use of the F-score” (pp. 37-38). Pavlick and Tetreault (2016) analyzed human perceptions of formality and their results also accorded with definition of context-dependence by Heylighen and Dewaele (1999). By reviewing linguistics literature, Pavlick and Tetreault (2016) further found that treating formality as a continuum is the supported view in this field. Larsson and Kaatari (2020) adhere to this view by stating that formality should be treated “as a cline rather than a dichotomy” (p. 12). In the same vein, by calculating the F-score, instead of a 0 or 100 distinction of formal versus informal, a relative score between the range of 0 to 100 is obtained.

Below are other characteristics of the F-score, stated by Heylighen and Dewaele (1999), which render it suitable for the present study:

1. F-score seems to be capable of doing its expected job. It distinguishes different types of language in terms of formality level as intuitively and theoretically expected by human judgment.
2. This formula is applicable to different samples or styles.
3. More fine-grained analyses are hardly more informative than more general ones. Therefore, considering the amount of time and effort spent on more fine-grained designs, working with more coarse-grained measures seem to be more efficient.

Furthermore, since assessing formality of smaller units of language (at sentence level) is more subjective (Lahiri et al., 2011; Hyland & Jiang, 2017), adopting a measure which captures a bigger picture of text formality seems to be more suitable for the goals of the present study.

Investigation of the Research Question: Finding the Degrees of Formality

The research question investigated the degree of formality in the native and non-native sample. F-score ranges from 0 to 100; a higher F-score reflects a higher level of formality (Heylighen & Dewaele, 1999). As it can be seen in Table 2, the native sample obtained a mean score of 50.27 and the non-native sample a mean score of 50.29 on a scale of 0 to 100. These numbers, according to Heylighen and Dewaele (1999), indicate a moderate amount of formality. Therefore, it can be concluded that English articles in the field of applied linguistics written by native and Iranian non-native researchers have a moderate amount of formality.

Table 2*Descriptive Statistics*

	Nativeness	N	Mean	Std. Deviation	Std. Error Mean
Formality	Native	40	50.271	.034	.005
	Non-native	40	50.291	.022	.003

Investigation of the Research Question: Comparing the Degrees of Formality

The research question was also posed to find out whether the native and non-native samples of this study have significantly different degrees of formality. First, in order to establish whether the dependent variable is normally distributed for each level of the independent variable, normality tests were done using SPSS.

Table 3*Tests of Normality*

	Nativeness	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Formality	Native	.110	40	.200*	.947	40	.060*
	Non-native	.117	40	.183*	.963	40	.216*

* $p < .05$

As it can be seen in Table 3, both tests show that the F-scores in both native and non-native groups are distributed normally ($p > .05$).

Consequently, in order to compare the means of both groups, a *t*-test for independent samples was calculated using SPSS (Table 4).

Table 4*Independent Samples t-Test*

		<i>t</i> -test for Equality of Means				
		t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference
Formality	Equal variances not assumed	2.984	66.968	.004*	.019	.006

* $p < .05$

The results of the *t*-test reject the null hypothesis of equality of means in the two groups and show that compared to native researchers, Iranian non-native authors of English utilized significantly higher degrees of formality in their articles, $t(66.968) = 2.984, p = 0.004$ at the specified level of $p < .05$, two-tailed, 95% CI [.032, .006].

In order to find the strength of the difference, Cohen's *d* was calculated ($d = 0.67$). This number, according to the guideline from Cohen (1988), indicates that the non-native group mean is higher than the native group by a medium effect size (of 0.67).

Discussion

The results of the statistical procedures show that though both native and non-native writers gained a moderate formality score, they are significantly different. Iranian non-native researchers obtained a higher mean score which, based on the *t*-test results, indicates that they write more formally in comparison with native researchers. These results are consistent with those of Alipour and Nooreddinmoosa (2018) and Ebrahimi and Fakheri (2019). They compared their results with those of Hyland and Jiang (2017) and concluded that applied linguistic articles published by Iranian researchers are more formal than those published by native writers. For example, in Ebrahimi and Fakheri (2019), contractions, which are normally considered as informal, were totally ignored by Iranian writers while this was not the case in non-Iranian sample in Hyland and Jiang (2017). What makes this study different is that both native and non-native samples were selected from the same international journals, whereas the Iranian non-native sample of both aforementioned studies are taken from articles published in Iranian local journals.

Possible Explanations for the Observed Difference

According to Ebrahimi and Fakheri (2019), this higher degree of formality can be due to two reasons. One is teaching methods of English in Iran which present academic writing as a strict inflexible genre in which writers are merely reporting the experiment and its results. The other reason is that Iranian postgraduate students are not required to write creatively and critically. As a result, fearing the rejection of their articles because of using an informal language, Iranian researchers see themselves as mere reporters of experiments. However, by doing so, they neglect the need for claims to be objective and representative of the writer-reader relationship. Martinez (2018) focused on first person distribution in different sections of biology papers produced by native and non-native speakers of English. Among other differences, it was found that native speakers used first person in a higher degree in the Results section to show that they assumed responsibility for their results. In addition, Ebrahimi and Fakheri (2019) mention that by merely reporting the experiment, Iranian authors are ignoring the fact that the writers' stance on the results is also valued by many journal editors when assessing articles for publication. Lack of such an authoritative stance, could be one of the reasons that some applied linguistics articles get rejected when submitted to well-known international journals. The following examples taken from the corpus can further illustrate this point.

Excerpt taken from the native sample:

“We are not reporting the results of a controlled experiment but rather an initial attempt to develop ...” (Poehner & Lantolf, 2013, p. 325).

“We therefore agree with Kozulin and Garb ...” (Poehner & Lantolf, 2013, p. 338)

Excerpt taken from the non-native sample:

“The current study suggests that ...” (Rahmati et al., 2019, p. 471)

“This study documented salient findings ...” (Tajeddin & Moghadam, 2012, p. 365)

In these excerpts, based on the above-mentioned explanations, the language of native researchers helped them take the responsibility of their research, while Iranian writers' language distanced them from their results and from taking the responsibility of arguments.

Native and non-native speakers of a language are also believed to have different levels of proficiency in different language skills. Lack of language proficiency can be stated as another probable source of the observed difference since expressing ideas in an informal yet acceptable language may be difficult for Iranian writers when writing in English. For example, research has shown that compared to native speakers, non-native speakers “often have a restricted range of lexis” (Hinkel, 2003, p. 294) and exhibit “a more restricted repertoire of recurrent word combinations” (Adel & Erman, 2012, p. 90). In the extreme case, it can be said that not having the proficiency required for greater specificity, non-native writers may end up using a term which sounds either too formal/informal in a specific context.

Incorporating an appropriate use of informality into academic writing could add to the complexity of this demanding task (e.g., Hyland & Jiang, 2017) particularly for those writing in a second language. Chang and Swales (1999) who took their sample from advanced writing classes for non-native speakers of English found that majority of their sample were uneasy about the increased use of informality and believed that it can make academic writing even more complicated.

There is no doubt that what is acceptable in a specific language depends on its social, cultural, and discursive conventions which, as mentioned by Johns (1990), may not be obvious to outsiders of a discourse community. Non-native English writers who do not possess a comprehensive knowledge and experience of these areas are faced with problems when writing in English.

Askarzadeh Torghabeh (2007) mentioned another problem that is rooted in the theory of transfer. It happens when learners of a new language transfer patterns of their native language and culture to that new language. Specifically, Iranian writers may transfer some linguistic or cultural patterns from academic writing in Farsi, which is believed to strictly follow conventions of formality, to English. Abdi (2010) compared the use of metadiscourse in English and Persian articles on sociology, education, psychology, physics, chemistry, and medicine. He found that the biggest difference between the two languages is in use of self-mentions, with Persian writers tending to write more impersonally than English authors. Also, Taki and Jafarpour (2012) studied sociology and chemistry articles in Farsi and English. They found that while Persian writers usually employ words like ‘the researcher’ to refer to

themselves, English writers use first person pronouns and possessive adjectives. This is exemplified below by a few excerpts taken from the corpus.

Excerpt taken from the native sample:

“Many students confided in me during interviews ...” (Gleason, 2014, p. 128)

“Results of our investigation suggest that ...” (Baecher et al., 2014, p. 131)

“If we did that, then the normal citation patterns for work in 2006 would be disrupted ...” (Meara, 2012, p. 9)

Excerpt taken from the non-native sample:

“The responses were coded and interpreted by both researchers ...” (Ghahari & Sedaghat, 2018, p. 12)

In the native excerpts, based on the above-mentioned explanations, the researchers freely talked about themselves by using personal pronouns. However, Iranian researchers in the non-native excerpt preferred an impersonal way of addressing themselves.

Faghih and Rahimpour (2009) also found that the type of metadiscourse Iranian researchers use in their articles on applied linguistics are different when they write in Farsi and English. Interestingly, they discovered a contradictory pattern of self-mentions being more frequently used in Farsi. Such observations can suggest that the rigidity Iranian authors presume to exist in English formal writing is not transferred from Farsi, but rather is probably due to a fixed framework they have in mind for English academic writing.

Durrant and Schmitt (2009) talked about non-native speakers' conservative approach to writing which manifests itself in usage of common forms of language and repetition of favored items. This can be seen as another reason for the observed difference between native and non-native samples. Having studied the attitudes of two French full professors who had their English articles rejected, Sionis (1995) revealed that they just wanted to learn "a few recipes and tips" to improve their skills in writing and just by doing that become successful writers who publish in English-medium journals (p. 100). Another possible explanation, therefore, is that non-native writers tend to view academic writing as a rather fixed style and follow its rules more strictly. In an Iranian context, this tendency can be attributed to "the more-or-less traditional type of instruction supplied in research writing and EAP classes in the Iranian academic context, which places an overemphasis on formal aspects of academic writing, overriding the role that informality can play in article development to engage readers" (Alipour & Nooreddinmoosa, 2018, p. 367). How native researchers are different from Iranian researchers in this aspect is exemplified by a few extracts from the corpus below.

Excerpt taken from the native sample:

“However, in 2009, we saw the publication of the first full-length volumes dealing with ...” (Forman, 2012, p. 239).

“In Table 5, we see an example of this process ...” (Forman, 2012, p. 246).

Excerpt taken from the non-native sample:

“In line with Walsh, it can be seen that Managerial Mode occurs most

often at the beginning of ..." (Ghafarpour, 2017, p. 215).

In these excerpts, based on the above-mentioned explanations, it can be seen that native researchers tried to engage their readers in their text by creating a kind of writer-reader community. Such an approach, however, cannot be found in the non-native excerpt.

Another possible explanation for the observed results may be attributed to the nature of applied linguistics papers which, as Hyland and Jiang (2017) observed, are less informal compared to those on biology, engineering, and sociology. For example, in their study of writer's presence in chemistry research articles written in English, Behnam et al. (2014) found that non-native Iranian writers use first person pronouns more than native ones. However, it should be noted that a feature like self-mention is just one among many other elements which affect formality of language.

Shift Toward Informality

In their book on teaching academic writing, Coffin et al. (2003) defined formality as the "use of technical, elevated or abstract vocabulary, complex sentence structures and the avoidance of the personal voice" (p. 28) and stated that "in particular we look at ways of avoiding an overly personal and anecdotal style when drawing on professional experience" (p. 68).

As it can be seen, formal speech has been treated as a "very correct and serious rather than relaxed and friendly" language (www.collinsdictionary.com, 2020) and informality has been defined as "the absence of full grammatical sentences, a decrease in concern about punctuation, and a high tolerance for typographic and spelling errors" (Coffin et al., 2003, p. 141). However, as it is discussed in the following paragraphs, what seems to be happening in academic writing right now is different from these definitions.

Bennett (2009) studied style manuals and revealed that some authors expect the grammar and lexis of academic writing to be formal while some other prefer "simplicity of style" (p. 52). Fairclough (2015) proposes the concept of 'synthetic personalization' which values being interpersonally involved in discourse. It roughly means personalization of the author and addressing the audience directly as individuals. Synthetic personalization is becoming increasingly widespread these days; therefore, it may not be easy to prevent even the most objective styles such as academic writing, which is believed to possess an air of detachment, from following this trend. In the same vein, Seone and Loureiro-Porto (2005) claim that in scientific English papers, passive is no longer the default choice, but actives are progressively replacing them. Nonetheless, the language of these papers does not seem to be undergoing colloquialization.

Though conventions of formality which necessitate avoiding informal elements are becoming less strictly followed compared to earlier times (e.g., Constantinou et al., 2020), as Swales and Feak (2012) state, "a formal research report written in informal, conversational English may be considered too simplistic, even if the actual ideas and/or data are complex" (p. 14).

To briefly describe this shift over the last decade or so, it can be said that "academic writing has gradually lost its traditional tag as an objective, faceless and impersonal form of discourse and come to be seen as a persuasive

endeavor involving interaction between writer and readers” (Hyland, 2005, p. 2). Informality should be regarded as the instantiation of a reader-friendly approach to writing which has expanded to writing academic texts (Hyland & Jiang, 2017).

It is worth mentioning that increase or decrease of informal features in different genres does not follow a fixed pattern. For example, Hyland and Jiang (2017) found that while second person pronouns/determiners had increased in applied linguistics papers, they had diminished in sociology articles.

Conclusion

The present study investigated the degree of formality in applied linguistics articles written in English by native and Iranian non-native researchers in international journals. It was found that native speakers of English write less formally than Iranian non-native researchers when writing in English.

Nowson et al. (2005) asserted that “there is little variation in the levels of formality per genre” (p. 33). Hence, it is worth mentioning here that though the degree of formality in two samples were significantly different, both of them, having come from the same genre, had a mean F-score of about 50 which shows a moderate amount of formality on the formality scale according to Heylighen and Dewaele (1999).

Types of instruction used in Iran for teaching English in general and for academic writing in specific, inadequate language proficiency, lack of social, cultural and discursive knowledge of English, transferring patterns of academic writing from Farsi to English, conservatism in following linguistic conventions, and nature of applied linguistics field which make incorporating informal voice an added burden in academic writing were mentioned as possible explanations for the observed result.

Distinctions between formal and informal language should not only be introduced but also be practiced. Corpus-based instruction can help non-native writers discover the existing patterns which may sound informal to them but are actually accepted and are being used by research communities. What is worth mentioning here is that this approach should not lead to mere repetition of what has been practiced and kill the creativity of individual writers. In other words, rather than being restricted by the so-called standard norms, non-native researchers would feel more comfortable and liberated expressing their own style of writing.

Research article writing is and should be treated as a specialized form of writing which requires specialized knowledge and training. The same is true about linguistic formality, that is, it needs specialized introduction and practice.

In addition to teachers of English in different areas (such as English as a second/foreign language, English for academic purposes, and language skills in general), journal reviewers, course designers and materials developers should also take formality and its situational flexibility into consideration to help learners, apprentice writers, novice researchers, and non-native speakers of English in general.

Academic writing in different disciplines is getting less formal (Adel, 2008; Hyland & Jiang, 2017; McCrostie, 2008) and all researchers need to be

aware of this shift. Research articles should be responsive to the demands of the new conditions. If academic writing is undergoing a gradual shift away from rigid detached styles to ones that encourage more personal involvement, all researchers need to be aware of this trend in order to be successful in the new publication environment. In this way, non-native researchers can improve their work by appropriate incorporation of informal elements into their writings just as expert native writers do. The same issues may suggest the emergence of a new style of academic writing which is less concerned with formality in general.

The limitation of this study was inherent to the concept of nativeness. Richards and Schmidt (2010) summarize characteristics of a native speaker as “a person who learns a language as a child and continues to use it fluently as a dominant language”. They further maintain that native speakers of a language “identify with a community where it is spoken” and possess “clear intuitions about what is considered grammatical or ungrammatical in the language” (p. 386). As abstract concepts, like identification with a language, emerge as characteristics of native speakers of that language (Liaw, 2004), deciding exactly who a native speaker is becomes a more complex task. In the same vein, preparing a sample in which native and non-native researchers could be fully distinguished in a clear-cut manner was not possible in the present study. While collecting the corpus, the present researcher decided on the native and non-native samples by checking the affiliations of the authors and institutions from which the articles were submitted.

Further studies can investigate whether other characteristics of writers other than nativeness, such as personality, background knowledge and previous experience in academic writing, play a role in determining the formality of their writings. In addition, more detailed analyses as to exactly what linguistic factors are contributing to the differences in formality can be conducted in future research.

References

- Abdi, R. (2010). Projecting cultural identity through metadiscourse marking: A comparison of Persian and English research articles. *Journal of English Language Teaching and Learning*, 1(212), 1-15.
https://elt.tabrizu.ac.ir/article_638.html
- Abu Sheikha, F., & Inkpen, D. (2010). Automatic classification of documents by formality. *Proceedings of the 6th International Conference on Natural Language Processing and Knowledge Engineering*. <https://doi.org/10.1109/nlpke.2010.5587767>
- Adel, A., & Erman, B. (2012). Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31(2), 81-92. <https://doi.org/10.1016/j.esp.2011.08.004>
- Alipour, M., & Nooreddinmoosa, M. (2018). Informality in applied linguistics research articles: Comparing native and non-native writings. *Eurasian Journal of Applied Linguistics*, 4(2), 349-373. <https://doi.org/10.32601/ejal.464196>
- Askarzadeh Torghabeh, R. (2007). EIL, variations and the native speaker's model. *Asian EFL Journal*, 9(4), 67-76.
- Baecher, L., Farnsworth, T., & Ediger, A. (2014). The challenges of planning language objectives in content-based ESL instruction. *Language Teaching Research*, 18(1), 118-136. <https://doi.org/10.1177/1362168813505381>
- Behnam, B., Mirzapour, F., & Mozaheb, M. A. (2014). Writer's presence in English native and non-native speaker research articles. *Procedia-Social and Behavioral Sciences*, 98, 369-374. <https://doi.org/10.1016/j.sbspro.2014.03.428>
- Bei, G. X. (2011). Formality in second language discourse: Measurement and performance. *Interdisciplinary Humanities*, 28(1), 32-41.
<https://www.researchgate.net/publication/324014572>
- Belcher, D. D., & Connor, U. (Eds.). (2001). *Reflections on multiliterate lives*. Multilingual Matters.
- Bennett, K. (2009). English academic style manuals: A survey. *Journal of English for Academic Purposes*, 8(1), 43-54. <https://doi.org/10.1016/j.jeap.2008.12.003>
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D. (1995). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge University Press.
- Biber, D., & Conrad, S. (Eds.) (2019). *Register, genre, and style* (2nd ed.). Cambridge University Press.
- Biber, D., Davies, M., Jones, J. K., & Tracy-Ventura, N. (2006). Spoken and written register variation in Spanish: A multi-dimensional analysis. *Corpora*, 1(1), 1-7.
<https://doi.org/10.3366/cor.2006.1.1.1>
- Biber, D., & Hared, M. (1992). Dimensions of register variation in Somali. *Language Variation and Change*, 4(1), 41-75.
<https://doi.org/10.1017/S095439450000065X>
- Canagarajah, A. S. (1996). "Nondiscursive" requirements in academic publishing, material resources of periphery scholars, and the politics of knowledge production. *Written Communication*, 13(4), 435-472.
<https://doi.org/10.1177/0741088396013004001>
- Canagarajah, A. S. (2002). *A geopolitics of academic writing*. University of Pittsburgh Press. <https://doi.org/10.2307/j.ctt5hjn6c>
- Celik, S. (2006). A concise examination of the artificial battle between native and non-native speaker teachers of English in Turkey. *Kastamonu Education Journal*, 14(2), 371-376.
<https://dergipark.org.tr/en/pub/kefdergi/issue/49104/626612>
- Chambers, J. K., Trudgill, P., & Schilling-Estes, N. (Eds.) (2003). *The handbook of language variation and change*. Blackwell Publishers.
- Chang, Y. Y., & Swales, J. (1999). Informal elements in English academic writing: Threats

- or opportunities for advanced non-native speakers? In C. Candlin & K. Hyland (Eds.), *Writing: Texts, processes and practices* (pp. 145-167). Longman.
- Coffin, C., Curry, M., Goodman, S., Hewings, A., Lillis, T., & Swann, J. (2003). *Teaching academic writing: A toolkit for higher education*. Routledge.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Conrad, S., & Biber, D. (Eds.) (2001). *Variation in English: Multi-dimensional studies*. Routledge.
- Constantinou, F., Chambers, L., Zanini, N., & Klir, N. (2020): A diachronic perspective on formality in students' writing: empirical findings from the UK. *Language, Culture and Curriculum*, 33(1), 66-83.
<https://doi.org/10.1080/07908318.2018.1563609>
- Curry, M. J., & Lillis, T. (2004). Multilingual scholars and the imperative to publish in English: Negotiating interests, demands, and rewards. *TESOL Quarterly*, 38(4), 663-688. <https://doi.org/10.2307/3588284>
- Dewaele, J.-M. (1996). How to measure formality of speech? A model of synchronic variation. In K. Sajavaara & C. Fairweather (Eds.), *Approaches to second language acquisition* (pp. 119-133).
- Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *International Review of Applied Linguistics in Language Teaching*, 47(2). <https://doi.org/10.1515/iral.2009.007>
- Ebrahimi, S. F., & Fakheri, S. A. (2019). Features of informality in applied linguistics research articles published in Iranian local journals. *Romanian Journal of English Studies*, 16(1), 135-143. <https://doi.org/10.1515/RJES-2019-0017>
- Faghih, E., & Rahimpour, S. (2009). Contrastive rhetoric of English and Persian written texts: Metadiscourse in applied linguistics research articles. *Rice Working Papers in Linguistics*, 1, 92-107.
- Fairclough, N. (2015). *Language and power* (3rd ed.). Routledge.
- Florence Ma, L. P. (2012). Advantages and disadvantages of native- and nonnative-English-speaking teachers: Student perceptions in Hong Kong. *TESOL Quarterly*, 46(2), 280-305. <https://doi.org/10.1002/tesq.21>
- Flowerdew, J. (1999). Writing for scholarly publication in English: The case of Hong Kong. *Journal of Second Language Writing*, 8(2), 123-145.
[https://doi.org/10.1016/S1060-3743\(99\)80125-8](https://doi.org/10.1016/S1060-3743(99)80125-8)
- Formal. (2020). In *Collins English Dictionary*.
<https://www.collinsdictionary.com/dictionary/english/formal>
- Forman, R. (2012). Six functions of bilingual EFL teacher talk: Animating, translating, explaining, creating, prompting and dialoguing. *RELC Journal*, 43(2), 239-253. <https://doi.org/10.1177/0033688212449938>
- Ghafarpour, H. (2017). Classroom conversation analysis and critical reflective practice: Self-evaluation of teacher talk framework in focus. *RELC Journal*, 48(2), 210-225. <https://doi.org/10.1177/0033688216631173>
- Ghahari, S., & Sedaghat, M. (2018). Optimal feedback structure and interactional pattern in formative peer practices: Students' beliefs. *System*, 74, 9-20.
<https://doi.org/10.1016/j.system.2018.02.003>
- Gleason, J. (2014). "It helps me get closer to their writing experience" Classroom ethnography and the role of technology in third-year FL courses. *System*, 47, 125-138. <https://doi.org/10.1016/j.system.2014.09.023>
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *Elementary School Journal*, 115(2), 210-229.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223-

234. <https://doi.org/10.3102/0013189X11413260>
- Heylighen, F. (1999). Advantages and limitations of formal expression. *Foundations of Science*, 4, 25-56. <https://doi.org/10.1023/A:1009686703349>
- Heylighen, F., & Dewaele, J. M. (1999). *Formality of language: Definition, measurement and behavioral determinants* (Internal Report). <http://pespmc1.vub.ac.be/Papers/Formality.pdf>
- Heylighen, F., & Dewaele, J. M. (2002). Variation in the contextuality of language: An empirical measure. *Foundations of Science*, 7, 293-340. <https://doi.org/10.1023/A:1019661126744>
- Hinkel, E. (2003). Simplicity without elegance: Features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37(2), 275-301.
- Hryniuk, K. (2015). Linguistics research articles written in English: Comparing native English speakers and Polish writers. *International Journal of Applied Linguistics*, 27(1), 3-23. <https://doi.org/10.1111/ijal.12099>
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Hyland, K., & Jiang, F. K. (2017). Is academic writing becoming more informal? *English for Specific Purposes*, 45, 40-51.
- Jenkins, J. (2003). *World Englishes: A resource book for students* (2nd ed.). Routledge.
- Jenkins, J. (2006). The spread of EIL: A testing time for testers. *ELT Journal*, 60(1), 42-50. <https://doi.org/10.1093/elt/cci080>
- Jenkins, J. (2011). Accommodating (to) ELF in the international university. *Journal of Pragmatics*, 43(4), 926-936. <https://doi.org/10.1016/j.pragma.2010.05.011>
- Johns, A. (1990). L1 composition theories: Implications for developing theories of L2 composition. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom*. Cambridge University Press.
- Kim, Y. J., & Biber, D. (1994). A corpus-based analysis of register variation in Korean. In D. Biber & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 157-181). Oxford University Press.
- Lahiri, S., Mitra, P., & Lu, X. (2011). Informality judgment at sentence level and experiments with formality score. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (pp. 456-457). Springer. https://doi.org/10.1007/978-3-642-19437-5_37
- Larsson, T., & Kaatari, H. (2020). Syntactic complexity across registers: Investigating (in)formality in second-language writing. *Journal of English for Academic Purposes*, 45(1), 1-16. <https://doi.org/10.1016/j.jeap.2020.100850>
- Liardét, C. L., Black, S., & Bardetta, V. S. (2019). Defining formality: Adapting to the abstract demands of academic discourse. *Journal of English for Academic Purposes*, 38(2), 146-158. <https://doi.org/10.1016/j.jeap.2019.02.007>
- Liaw, E. (2004). "How are they different?" *A comparative study of native and nonnative foreign language teaching assistants regarding selected characteristics: Teacher efficacy, approach to language teaching/learning, teaching strategies and perception of nativeship* [Unpublished doctoral dissertation]. The Ohio State University, Ohio, USA. <https://etd.ohiolink.edu/>
- Lillis, T. (2013). *The sociolinguistics of writing*. Edinburgh University Press.
- Lillis, T., & Curry, M. J. (2006). Professional academic writing by multilingual scholars: Interactions with literacy brokers in the production of English-medium texts. *Written Communication*, 23(1), 3-35. <https://doi.org/10.1177/0741088305283754>
- Lillis, T. & Curry, M. J. (2010) *Academic writing in a global context: The politics and practices of publishing in English*. Routledge.
- Martinez, R. (2018). "Specially in the last years...": Evidence of ELF and non-native English forms in international journals. *Journal of English for Academic Purposes*, 33, 40-52. <https://doi.org/10.1016/j.jeap.2018.01.007>

- McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal*, 32, 97-114. http://icame.uib.no/ij32/ij32_97_114.pdf
- Meara, P. (2012). The bibliometrics of vocabulary acquisition: An exploratory study. *RELC Journal*, 43(1), 7-22. <https://doi.org/10.1177/0033688212439339>
- Mesthrie, R., Swann, J., Deumert, A., & Leap, W. L. (2009). *Introducing sociolinguistics* (2nd ed.). Edinburgh University Press.
- Moreno, A. I. (2010). Researching into English for research publication purposes from an applied intercultural perspective. In M. F. Ruiz-Garrido, J. C. Palmer-Silveria, & I. Fortanet-Gómez (Eds.), *English for professional and academic purposes* (pp. 57-71). Rodopi.
- Mur-Duenas, P., & Jolanta, Š. (2016). Self-reference in research articles across Europe and Asia: A review of studies. *Brno Studies in English*, 42(1), 71-92. <https://doi.org/10.5817/BSE2016-1-4>
- Nowson, S., Oberlander, J., & Gill, A. J. (2005). Weblogs, genres, and individual differences. *Proceedings of the 27th Annual Conference of the Cognitive Science Society*, 1666-1671.
- Ozturk, U., & Atay, D. (2010). Challenges of being a non-native English teacher. *Educational Research*, 1(5), 135-139.
- Pavlick, E., & Tetreault, J. (2016). An empirical analysis of formality in online communication. *Transactions of the Association for Computational Linguistics*, 4(2), 61-74. https://doi.org/10.1162/tacl_a_00083
- Peterson, K., Hohensee, M., & Xia, F. (2011). Email formality in the workplace: A case study on the Enron corpus. *Proceedings of the Workshop on Languages in Social Media*, 86-95. <https://www.researchgate.net/publication/262311292>
- Poehner, M. E., & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during Computerized Dynamic Assessment (C-DA). *Language Teaching Research*, 17(3), 323-342. <https://doi.org/10.1177/1362168813482935>
- Rahmati, T., Sadeghi, K., & Ghaderi, F. (2019). English language teachers' vision and motivation: Possible selves and activity theory perspectives. *RELC Journal*, 50(3), 457-474. <https://doi.org/10.1177/0033688218777321>
- Richards, J. C., & Schmidt, R. W. (2010). *Longman dictionary of language teaching and applied linguistics* (4th ed.). Pearson.
- Sardinha, T. B., & Pinto, M. V. (Eds.). (2014). *Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber*. John Benjamins.
- Seone, E., & Loureiro-Porto, L. (2005). On the colloquialization of scientific British and American English. *ESP Across Cultures*, 2, 106-118.
- Sionis, C. (1995). Communication strategies in the writing of scientific research articles by non-native users of English. *English for Specific Purposes*, 14(2), 99-113.
- Strauss, P. (2017). "It's not the way we use English": Can we resist the native speaker stranglehold on academic publications? *Publications*, 5(4), 1-7. <https://doi/10.3390/publications5040027>
- Strauss, P. (2019). Shakespeare and the English poets: The influence of native speaking English reviewers on the acceptance of journal articles. *Publications*, 7(1), 1-10. <https://doi:10.3390/publications7010020>
- Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: A course for nonnative speakers of English* (3rd ed.). Ann Arbor, Michigan: University of Michigan Press.
- Tajeddin, Z., & Moghadam, A. Z. (2012). Interlanguage pragmatic motivation: Its construct and impact on speech act production. *RELC Journal*, 43(3), 353-372. <https://doi/10.1177/0033688212468481>
- Taki, S., & Jafarpour, F. (2012). Engagement and stance in academic writing: A study of English and Persian research articles. *Mediterranean Journal of Social Sciences*

3(1), 157-168. <https://www.researchgate.net/publication/236026941>

Tardy, C. (2004). The role of English in scientific communication: Lingua franca or Tyrannosaurus rex? *Journal of English for Academic Purposes*, 3(3), 247-269. <https://doi.org/10.1016/j.jeap.2003.10.001>



©2020 Alzahra University, Tehran, Iran. This article is an open-access article distributed under the terms and conditions of the Creative Commons [Attribution-NonCommercial 4.0 International \(CC BY-NC-ND 4.0 license\)](https://creativecommons.org/licenses/by-nc-nd/4.0/) (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)