Referential Coherence of Academic Texts: A Corpus-Based Analysis of L2 Research Papers in Management

Elizaveta A. Smirnova

National Research University Higher School of Economics

Correspondence concerning this article should be addressed to Elizaveta A. Smirnova, Associate Professor of Foreign Languages Department, National Research University Higher School of Economics, 38 Studencheskaya St., Perm, 614070, Russian Federation. E-mail: esmirnova2@hse.ru

This paper focuses on referential coherence, which is seen as a crucial attribute of effective academic writing. Findings are reported from a corpus study of Russian students' research proposals. The learners' use of anaphoric expressions is compared with a reference corpus, which comprises research articles published in peer-reviewed journals. It was hypothesised that learners use anaphora less frequently than professional writers and face some difficulties when using anaphoric expressions. The results of the analysis partly confirmed the hypothesis and allowed the identification of particular problems connected with the students' use of anaphoric expressions, which were then classified into several groups. Examples of exercises aimed at dealing with the identified problems are also provided. It is hoped that the reported findings, as well as the author's suggested reasons for the problems and possible ways of dealing with them, will be useful for EAP practitioners, researchers, and students writing their research papers in English.

Keywords: referential coherence; second language writing; corpus analysis; anaphora; EAP

Introduction

Despite the existence of a substantial number of studies on text coherence, this significant attribute of academic style remains under-researched in the field of L2 writing. Coherence in an academic text is no less important than its content, however, it frequently turns out to be difficult to achieve, especially for second language learners (see, for example, Al Sharoufi, 2014; Hyland, 2004; Hyland & Tse, 2004; Leńko-Szimańska, 2004), therefore demanding the close attention of researchers and teachers of academic writing.

Coherence is researched in two aspects: *relational* and *referential* (see, for example, Degand et al., 1999). Relational coherence and *coherence relations* are referred to as "the cornerstone of comprehension" (Graesser et al., 2003, p. 82). They comprise, for instance, List, Cause and Consequence, and Problem and Solution, which are expressed in a text by various linguistic markers, such as connectives and lexical cue phrases. The other type of coherence, called referential coherence, is achieved via the use of referential markers, mainly *anaphora*. Anaphora is defined as a language phenomenon that introduces the connection between a "pointing back expression", called *anaphor* (or *referent*), and *antecedent*, i.e. the text segment an anaphor refers to (see, for example, Charaudeau & Maingueneau, 2000, p. 45; p. 14; Ducrot & Schaeffer, 1995: p. 458; Halliday & Hasan, 1976; Matysiak, 2007, p. 183). Anaphora plays an important role in a text because it relates the current information to the intended referent. This is essential for expressing the thematic structure of the discourse and its comprehension (Vonk et al., 1992).

EAP learners are generally familiar with relational coherence because it is frequently addressed in various textbooks and guides on academic writing, some of which even provide lists of linking expressions to be used in different situations.[1] However, referential coherence markers, i.e. anaphora, are rarely focused on in EAP courses,

---

[1] Bailey, S. (2006). Academic Writing: A handbook for international students. New York, NY: Routledge, UK; Greetham, B. (2001). How to write better essays. New Yourk, NY: Palgrave; Mounsey, C. (2002). Essays and dissertations. Oxford, UK: Oxford University Press; Soles. D., & Lawler, G. (2005). The academic essay: How to plan, draft, write and revise. Abergale, UK: Glmp Ltd.

despite the fact that the use of anaphoric expressions deserves close attention, being a crucial component of academic text coherence. Anaphora frequently occurs in academic texts, which makes it an essential topic for academic writing instruction (Crompton, 2017). It was assumed that L2 writers have even more difficulties than L1 writers with the effective use of anaphoric expressions because anaphora is quite a complex linguistic feature whose accurate use requires a high level of language proficiency.

This work aims at investigating anaphoric expressions as a marker of coherence in an academic text. Being an EAP practitioner who often deals with students' papers and basing on previous research findings in this field (e.g., Leńko-Szimańska, 2004), the author hypothesised that L2 learners underuse referential markers in their texts in comparison with expert writers. It was also expected to find some specific problems with the use of anaphora caused by the students' lack of English language proficiency and the influence of their native language. It was posited that researching the way L2 learners use anaphora in their texts and comparing it to experts' usage could allow some problems the students face when using anaphoric expressions to be detected. Understanding the reasons can direct teachers to more efficient ways of tackling the issues in EAP and ESL classrooms.

Therefore, it was assumed that the research findings might be useful for EAP practitioners, researchers, and students writing their research papers in English.

First, it is necessary to mention that the term *coherence* is often used alongside *cohesion*. Some scholars differentiate between them, stating that cohesion contributes to coherence, i.e. grammatical and lexical units used in a text to create connections between its parts signal coherence (see, for example, de Beaugrande, 1980; Halliday & Hasan, 1976; Tanskanen, 2006; Witte & Faigley, 1981). De Beaugrande and Dressler (1981) claim that coherence is "not a mere feature of texts, but rather the outcome of cognitive processes among text users" (De Beaugrande & Dressler, 1981, p. 6). In some works (see, for example, Halliday and Hasan, 1976)[2], the word 'cohesion' is used to refer to the grammatical and lexical markers used to make a text cohesive, while Witte and Faigly make a clear distinction between Halliday and Hasan's 'cohesion' and 'coherence' (1981, : p 199), the former limited to the words and syntax, the latter involving the broader social-cultural context. There are also some authors who do not distinguish between the two terms, such as Kleijn et al. (2019). Following Fahnestock (1983) and Degand et al. (1999), in this paper, the term *coherence* is used as a more succinct one to refer to the relations existing between successive sentences and paragraphs (Degand et al., 1999) formed in a text via various linguistic means.

Coherence in different types of writing has been explored by a number of scholars. For example, Bunton (1999) focused on textual signaling in PhD theses of Hong Kong research students. He concluded that despite some inconsistency in their use, overall, textual signals contribute to maintaining coherence of a long text. Hyland (2004) and Hyland and Tse (2004) analysed an extensive corpus of ESL learners' masters and doctoral dissertations. The studies showed a high frequency of metadiscoursal devices, namely transitions, which contribute to a text's coherence, in the works under consideration the frequency was higher in doctoral in comparison with masters dissertations. Some studies of coherence in students' writing showed English L2 students tend to over-use particular cohesive devices, for example, linking adverbials (Shaw, 2009) or use them differently from the way native speakers do (Gardezi & Nesi, 2009). Interactive and Interactional metadiscourse markers in Chinese students' academic writing in two disciplines, Literary Criticism and Translation Studies, were studied by Li and Wharton (2012). The featured students studied through the medium of English in different universities located in China and the UK. It was found that local institutional conventions had a stronger effect on the learners' use of metadiscourse markers than disciplinary factors.

Degand et al. (1999) explored *relational* and *referential* coherence, which are aimed at helping the reader integrate the text contents and discourse values into a global structure for easier interpretation. Both types of coherence act as tools for establishing "the continuity of meaning and reference", which make a mere sequence of clauses and sentences become a text (Cornish 2009, p. 572-573). However, while relational coherence has been paid substantial attention in linguistics literature (see, for example, Hoek et al., 2017; Kleijn, Pander Maat, & Sanders, 2019; Sanders, Land & Mulder, 2007; Sanders & Noordman, 2000) and EAP textbooks (see, for example, Bailey, 2006; Greetham, 2001; Mounsey, 2002; Northedge, 2005; Soles & Lawler, 2005), referential coherence appears to be understudied.

---

[2] Halliday, M. A., & Matthiessen, C. (2013). Halliday's Introduction to Functional Grammar (4th ed.). Milton Park, UK: Routledge.

Referential coherence, as mentioned above, is realized in a text by referential markers, namely anaphora (Degand et al., 1999). Anaphors are seen as "an essential pillar supporting the overlay of [coherence] relation in fleshing out and integrating their discourse values" (Cornish, 2009, p. 590). Several studies have posited that the absence of referential coherence between two successive sentences or clauses has a negative influence on comprehension (see, for example, Clark & Haviland, 1977; Lesgold et al., 1979). Degand et al. (1999) also studied the influence of linguistic markers of referential coherence on the comprehension of expository discourse. They prepared texts in which anaphors where difficult to resolve and asked respondents to read them and answer some questions to check comprehension. The researchers found that less coherent texts were harder to understand. The role of coherence markers on the 'readability' of an academic text was explored by Basturkmen & von Randow (2014) who observed the use of signals of text organisation and concession relations with and without signalling in postgraduate argumentative essays. The authors concluded that postgraduate students are generally well aware of the necessity to guide the reader via the use of 'reading cues'; however some of them might use those cues inappropriately.

The ability of second language writers to use anaphoric expressions in their texts has also been explored; however, the number of such studies is quite scarce. Thus, Plakans and Gebril (2017) studied anaphors alongside with logical operators as a measure of cohesion in TOEFL iBT writing tasks. They found that the effect size for anaphor reference was smaller than that for logic operators. Al Sharoufi (2014) analysed the use of anaphora, cataphora, and transitional signals in L2 students' academic essays. The author claims that the use of referential elements causes some difficulties for students, but the paper does not specify what particular problems with the use of anaphora were identified. Leńko-Szimańska (2004) focused on specific patterns of misuse of demonstratives as anaphora in Polish students' essays. She concluded that most problems occur with the use of *that* and *those* as determiners, which are overused by learners in comparison with the native speaker norm. Another finding was that Polish students tend to underuse *those* as a demonstrative pronoun. These errors are explained by the influence of the Polish language, which lacks articles.

Overall, earlier studies investigated the use cohesive devices in L2 students' writing and reported positive relations between cohesion features and writing quality. However, research into referential coherence in learners' texts seems to be very limited. Thus, the characteristics of use of anaphoric expressions by L2 learners have yet to be explored. Exploring referential markers could facilitate a better understanding of how the coherence of an academic text is achieved providing links to pedagogical interventions and writing strategy development.

The aforementioned studies either did not explicitly take into account the different linguistic means that can be used to establish this referential coherence, e.g., nominal, verbal, and adverbial anaphora, and their use in a text, or explored only a limited number of them (see, for example, Leńko-Szimańska, 2004). In this paper, the aim is to focus on three types of anaphora that contribute to the creation of coherence in an academic text and to research the use of anaphoric expressions in L2 students' writing by comparing them with examples found in experts' texts.

Materials

The analysis was based on two corpora. The learner corpus was a collection of 58 samples of students' writing. The writers were fourth-year management students of the National Research University Higher School of Economics, Perm, Russia. The samples were research proposals which they wrote as part of the course of English for Academic Purposes. The research proposals were brief descriptions of the students' diploma projects written in English amounting to roughly 2500 words. Russian is the native language of all the learners. Their level of proficiency in English is approximately B2-C1, according to the Common European Framework of Reference for Languages (CEFR). The learner corpus data were compared to a reference corpus.

The reference corpus comprises articles from peer-reviewed journals in Management – Journal of Management (JM) – 25 articles; Journal of Management Studies (JMS) – 20 articles; and Academy of Management Journal (AM) – 17 articles (62 in total). They are all leading academic journals with a current H index of 145, 119, and 227 (2017) correspondingly, which attests to them being quite influential in the international business and management studies discourse community. All publications refer to the period from 2013 to 2016. The full list of the sources quoted in this paper can be found in Appendix 2.

It was assumed that the two corpora were comparable since a paper published in a peer-reviewed journal is considered to be a model for students to follow, i.e. a perfect example of an academic text – a text type that comprises various research-oriented genres, such as dissertations, monographs, and presentations (Swales 1990). Therefore, it is necessary for EAP learners to conform to the conventions of the genre (Hiltunen, 2016), which include the use of anaphoric expressions. When studying the distinctive characteristics of L2 texts, it is also important to base the research on a norm (Granger, 2015). In the EAP context, this norm, as mentioned above, is taken to be represented by articles from peer-reviewed journals, therefore such comparisons appear to be very useful for pedagogical purposes. This practice is well-established in learner corpus research. For example, Lee and Chen (2009), who compared a corpus of expert academic writing consisting of journal articles and a corpus of undergraduate dissertations, claim that the comparison gave them "added confidence" that they were focusing on the right issues (p. 154).

The corpora amount to approximately 130,000 and 694,000 words respectively (after all tables, formulas, graphs, charts, and reference lists were removed from the texts). Since the study takes into account normalized frequency rates per 1,000 words, it is assumed that the differences in the size of corpora do not distort the overall tendencies identified during the analysis. The comparison of a sample corpus (a learner corpus in this case) to a larger corpus is also considered to be appropriate. In this case, the larger corpus represents a text norm against which the second corpus is compared (Rayson & Garside, 2000).

Another argument for the comparability of the corpora is that despite the fact that the texts the corpora consist of are from different genres, i.e. research proposals and research articles, they both belong to academic discourse and have very similar structures. They both describe the topic of the research and the gap in knowledge, include literature review on the topic, outline research methods, and give the results of the data analysis (anticipated results in the case of students' works).

Research Design

All texts were first processed with the help of the AntConc concordance programme, which is typically applied as a text analysis tool for processing corpus data. Nevertheless, it was also necessary to examine the texts manually in order to obtain both qualitative and quantitative results because some of the language means under consideration can perform different functions (e.g. *it* can be a subject pronoun and an object pronoun; *her* – an object pronoun and a possessive adjective). In a similar vein, some potential anaphora expressions could in fact not be performing an anaphoric function, e.g. *It* as in example 1. Therefore, the contexts of use of some tokens needed to be considered in order to select relevant examples. Thus, instances with *it* used as a dummy subject were not taken into account, while sentences with *it* used as a personal or object pronoun (examples 2 and 3) were included into the analysis.

(1)   ***It** is necessary to examine the degree to which the venture is structured in a manner that promotes free and open communication and knowledge sharing* (JMS 2016-8).

(2)   *When a new practice is introduced that employees do not support, **it** may lead employees to withhold their effort and commitment* (AM 2015-1).

(3)   *Our qualitative results show that such time before compliance is often invested in activities that help decision makers to learn more about the practice itself, about others' reactions to **it*** (AM 2015-1).

For qualitative analysis, contrastive interlanguage analysis (CIA) (Gilquin, 2000/2001; Granger, 1996) was employed to identify specific features of the use of anaphoric expressions in the L2 learner corpus. This methodology involves the juxtaposition of learner language and an expert speaker reference corpus in order to identify differences and similarities between them. Despite the fact that in the CIA framework the terms 'underuse' and 'overuse' are considered to be neutral and descriptive, not prescriptive (Granger, 2015), in terms of teaching EAP, the revealed frequency differences require close analysis in order to detect possible systematic problems students might have with the use of anaphoric expressions. The observed discrepancies in the learners' and expert writers' use of anaphora were then classified into four groups based on the types of errors students make.

The study focuses on three types of anaphora identified by Gardelle (2012): nominal (4), verbal (5), and adverbial (6) in order to get a more comprehensive picture of the use of anaphora in academic discourse.

(4) *Our experimental results revealed that **decision makers** saw opportunities for increased legitimacy if **they** complied early with the coercive demand primarily when **they** faced low institutional complexity. As institutional complexity increased, **they** planned increasing delays in **their** compliance* (AM 2015-1).

(5) *Hospitals that **did not offer robotic surgery**, especially when other nearby hospitals **did**, risked losing appeal and patients* (AM 2015-6).

(6) *I belong to that category of people that, I mean, if I read one night about a new intervention performed in **Boston**, the following day I was rushing **there** to see how they do it* (AM 2015-6).

The referential markers selected for analysis were: subject, object, reflexive and indefinite pronouns, demonstratives (both pronouns and adjectives), possessive adjectives, auxiliary verbs *do, does, did* functioning as substitutes, and the adverbs *here* and *there*.

Analysis

The significance of the differences in frequencies was calculated with log-likelihood tests (Dunning, 1993). According to Smith (2009), in corpus studies log-likelihood values can be interpreted in the following way: the log likelihood ⩾ 6.63 means that the result is significant at the 1% level ($p < 0.01$); the log likelihood ⩾ 3.84 means that the result is significant at the 5% level ($p < 0.05$).

In order to measure the size of the difference of normalised frequency of a word in the two corpora effect size metric (%DIFF) was used. It indicates the proportion of the difference. Negative %DIFF values show that the word has a higher normalized frequency in the reference corpus (Gabrielatos & Marchi, 2012) than in the leaner one.

Results

Nominal anaphora

The language units under consideration and the results of the corpus analysis are presented in Table 1.

Overall, nominal anaphora was marginally more frequently used by expert writers than by learners (21070 (30.360 per 1000 words) occurrences in total in the reference corpus vs 3865 (29.730 per 1000 words) occurrences in the learner corpus). In some cases, the log-likelihood values demonstrated that the difference between the learner and reference corpora was significant and worth attention.

Subject pronouns. Generally, learners used subject pronouns in their writing less frequently than expert writers. The log-likelihood values show the significance of the difference between the learner and expert corpora at the 1% level. The largest significant difference was detected for the subject pronoun *they* (%DIFF = -54.73).
The results of the corpus analysis show that the only subject pronoun students tend to use more often in comparison with expert writers was *it* (approximately 3.5 occurrences per 1000 words in the learner corpus vs about 2.3 in the reference corpus) which was the second largest significant difference among all the subject pronouns under consideration (%DIFF = 53.43).

Possessive adjectives. The largest effect size value was found for *its* (%DIFF=95.43), which alongside with *his* was more frequently used by learners. However, the difference of use for the latter was found to be insignificant, while the former was significant at the 1% level, demonstrating the second most significant log-likelihood value in the group of 96.22. *Their* and *her* were much more common in the reference corpus than in the learner one (%DIFF=-54.34 and -42.82 respectively).

Object pronouns. Despite the fact that the difference in the use of object pronouns between the two corpora turned out to be insignificant, there were some interesting findings. For example, as can be seen from the table, in the reference corpus, object pronouns were used more extensively with the largest effect size value of -67.65 found for *her*. The exception was *it* which prevailed in the learner corpus.

Reflexive pronouns. In professional writing reflexive, pronouns occurred more often than in student writing, with *itself* being an exception: it was used by learners more than twice as frequently as by expert writers and had the largest significant difference with %DIFF=109.73.

Demonstratives. The difference in the use of all the demonstratives between the learner corpus and the reference corpus was significant at the 1% level. Demonstratives in general were more common in the learner corpus (the normalized frequencies are 10.46 for the learner corpus and 9.31 for the reference one). This result is consistent with Leńko-Szimańska's findings (2004) who revealed that learners generally overuse demonstratives when compared with native speakers. Students tended to use *this* much more often than expert writers and this difference had the largest effect size value of 49.33.

Indefinite pronouns. The difference in the total use of possessive adjectives between learners and expert writers was the largest significant difference among all the anaphoric expressions analysed (%DIFF=69.57). The significant difference in the use of indefinite pronouns was found only for *one*. Interestingly, it was used more than three times as often in the learner corpus as in expert writing, while *ones* was more frequently employed in the reference corpus.

Relative pronouns. The largest effect size value in the group of relative pronouns was found for *who*, while the difference in the use of *whose* in the learner and reference corpora happened to be insignificant. *Who* was more often used by expert writers, whereas *which* occurred more frequently in the learner corpus.

Verbal anaphora

Halliday and Hasan (1976) posited that verbal anaphora is important for coherence because it serves to link sentences in a text in "the same way as the nominal substitute one" (p. 113). Nominal anaphora and verbal anaphora share many common features. For example, like the former, the latter has "an experiential structure in which the lexical verb expresses the 'Thing'", i.e. an action, event, or relation (Halliday & Hasan, 1976, p. 113).

According to the data in this study, verbal anaphora is a less frequent phenomenon in academic discourse than nominal anaphora. The results of the corpus analysis of verbal anaphora are presented in Table 2.

The data clearly show that verbal anaphora was much more common in the reference corpus that in the learner one. However, a significant difference was only detected for *do*.

Adverbial anaphora

The results of the analysis are presented in Table 3.
Table 3 shows that adverbial anaphora was more frequently used by expert writers than by students. However, the difference only appeared to be significant for *here*.

Discussion

The quantitative analysis of the referential coherence, which is realised by the three types of anaphora, showed that overall it is more common in expert writing than in students' papers, thus confirming the hypothesis. The differences in the use of anaphoric expressions in the two corpora were further explored through qualitative analysis. Problems in the learners' use of anaphoric expressions were classified into four groups and reasons for why the students wrote in a certain way will be discussed.

Repetition of a chunk of language (a word or phrase). The corpus examples reveal that the students tended to repeat nouns rather than omit repeating words or replace them with pronouns (21 cases in the corpus). This

might be responsible for the revealed underuse of nominal anaphora in the learner corpus. Thus, examples 7a and 7b demonstrate the sentences in which the subject pronouns *he* and *they,* which were found to be underused by the learners in comparison with expert writers, could have been employed instead of the repeated noun.

(7)   (a) ***Manager*** *eliminates the suppliers whose mark, with respect to a criterion, is lower than the minimal mark. Thus,* ***manager*** *chooses one of the suppliers satisfying the minimum level of all the criteria.*

(b) *Although* ***respondents*** *are likely to be understood and to allow that such goods will be more expensive, but for ensuring environmental safety* ***respondents*** *agree to pay.*

In examples 8a and 8b the use of the indefinite pronoun ones, which was also underused by learners, would have been appropriate, however the students preferred to repeat the plural noun:

(8)   (a) *Companies that appreciate and care about employees, are more likely to attract effective* ***employees***.

(b) *The first group of theories consists of utilitarian* ***theories***.

Similarly, the learners tended to avoid using auxiliary verbs as substitutes, which was also revealed by the qualitative analysis (see Table 2). They preferred to repeat the main verb(s), which results in longer sentences on the one hand, and bad style and a lack of coherence on the other:

(9)   *Many organizations strive to reduce production loss to gain an advantage over their main competitors or to reduce costs, but only few companies manage to* ***gain it and reduce their costs***.

Like in the cases of nominal and verbal anaphora, the findings showed that some students preferred to use repetition rather than employ adverbial anaphora to avoid it. Thus, in example 10, instead of repeating in the organisation twice, there could have been used in the second part of the sentence:

(10)   *Observation will include the observation of the current situation in the organization, complete examination of all the key processes* ***in the organization*** *and identification of the bottlenecks.*

The motivation for repeating words in the learners' texts might lie in the learners' not feeling confident about using anaphoric expressions, namely indefinite pronouns and verb substitutes, because they are rarely taught this in ESL courses. This goes in line with the findings of Ferris (1994) who investigated essays of L2 writers of different levels of language proficiency. She concluded that in order to achieve textual cohesion, lower proficiency writers tended to rely on repletion, while the more advanced learners employed a variety of referential cohesion devices.

To deal with the issue of learners' repeating chunks of language, teachers could offer students sentences with repetition taken from learner corpora for editing with the use of different kinds of anaphora. An example of such an editing exercise is presented in Appendix 1.

Superfluous use of anaphoric expressions. The analysis of concordance lines revealed that learners sometimes use anaphoric expressions unnecessarily (18 cases in the corpus), like in the case of *it* in 11a and 11b. This correlates with the results of the quantitative analysis showing that the students used the subject pronoun *it* more frequently than expert writers.

(11)   (a) *Furthermore, this method is the least resource-consuming and* ***it*** *allows to test the hypothesis on a real market, using MVP.*

(b) *The main advantages of SWOT are that it is simple to conduct and* ***it*** *is not costly for the organization.*

The analysis of corpus examples showed that in some writings the demonstrative *this* occurred as many as 44 times; and it also occurred more frequently in the learner corpus than in the expert one. That figure appears to

be large in comparison with the average use of *this* in the learner corpus, which accounted for 18 instances per text. Examples 10a and 10b demonstrate the learners' overuse of *this*:

(12) (a) *However, **this** paper supposes the use of customer development methodology for several reasons. Firstly, **this** method was adapted to IT-projects, for example, **this** technique involves the use of the approach of agile software development. Furthermore, **this** method is the least resource-consuming and it allows to test the hypothesis on a real market, using MVP. Finally, **this** technique was actually applied in practice in the majority of successful start-ups that participated in start-up accelerator of Russian Internet Initiatives Development Fund.*

(b) ***This** model is quite conventional. Although **this** approach is quite prejudiced, in any case it would be better to consider **this** approach.*

The overuse of *this* in the learner corpus might be due to the fact that the students tended to use *this* instead of the definite article or any other modifiers. This can be explained by the influence of their native language, where there are no articles. Therefore, the learners, being afraid of making an error with the definite article in English, preferred to use *this* to be on the safe side. This finding goes in line with Leńko-Szimańska's research (2004), which also explained Polish students' overuse of demonstrative pronouns by the influence of their native language.

It is the author's belief that providing students with extensive practice of the use of articles and closer attention to words and phrases that could replace demonstratives in a sentence may help to solve these problems. For example, in-class analysis of choices utilized by expert writers can help to improve learners' writing (see Exercise 2 in Appendix 1, which demonstrates the use of the definite article for establishing relational coherence in professional writing).

Plural and singular forms. The author observed some systematic errors that learners make when referring to plural nouns (15 cases in the corpus). This concerns almost all types of anaphoric expressions used to refer to plural nouns, namely the subject and object pronouns *it,* which are used instead of *they* and *them* correspondingly (see examples 13, 14a and 14b); as well as the possessive adjective *its* (example 15) and the demonstrative pronoun *this* (examples 16a and 16b). This finding also correlates with the results of the quantitative analysis, which revealed general underuse of plural forms of nominal anaphora in the learner corpus:

(13) *Methods are effective only if **it** brings results in accordance with the goals and objectives.*

(14) (a) *To obtain more specific information on each point I will select criteria and make a comparison of results according to **it**.*

(b) *The coach reflects the client's actions and helps to transform **it** into autonomous abilities (functions).*

(15) *The main idea in the third sub-group of corporate citizenship is that corporations can take **its** rightful place in society, next to other "citizens".*

(16) (a) *The interview questionnaire will be developed on the basis of the reviewed literature, and it will help in gathering all the relevant data from **this** major informants.*

(b) ***This** beliefs and expectations produce norms that powerfully shape the behavior of individuals and groups in the organization*

The misuse of plural and singular forms, namely demonstrative pronouns, cannot be explained by L1 interference, as the variability of demonstrative pronouns is typical of the Russian language (Rahimova & Yusupova, 2015). These findings suggest that plural forms, like articles, demand special attention in academic writing instruction. Once aware of the problem, an EAP teacher can anticipate and prevent possible mistakes in plural forms by drawing students' attention to the relevant grammar rules and provide examples of their use during the initial stages of the course. For example, error correction exercises focusing on plural forms of nominal anaphora can be used to solve the problem (see Exercise 3 in Appendix 1).

Wrong patterns. Examples of patterns used by the learners that were different from the norm, i.e. those used in the reference corpus, were found for possessive adjectives as well as reflexive and relative pronouns (19 cases in the corpus).

In the reference corpus the possessive adjective *their* was commonly used when referring to a person whose gender is unknown:

(17) (a) *As long as a person does **their** job, they can expect to stay in their job* (AM 2015-3).

(b) *In response, the employee regulates **their** emotion and emotional expression to meet the requirements of **their** job and achieve organizational goals* (JMS 2014-8).

However, in the learner corpus *his* was usually used in this situation:

(18) (a) *The method of coaching is based on the communicative cooperation in building partnerships between the coach and **his** client.*

(b) *The owner of business conducts market analysis when **he** communicates with **his** customers about business or checks the prices of **his** competitors.*

The use of *his* (as well as *he* in 18b) would have been justified if the gender of the coach from 16a or the business owner from 18b had been revealed earlier in the texts, but it was not the case. The learners were trying to make some general statements without referring to a particular person. This might be explained by the influence of Russian in which the noun 'person' is masculine, so it is common to use masculine pronouns to refer to a person whose gender is unknown. According to Pavlou and Potter (1994), Russian, being a highly inflected language, has difficulty responding to the societal demand of avoiding gender-biased language.

Another anaphoric expression that causes difficulties is the reflexive pronoun *themselves*. The most common patterns in which *themselves* was used in the reference corpus are shown in Table 4.

However, in the learner corpus the most common pattern with *themselves* (5 cases out of 12) was 'a verb + a noun + themselves':

(19) *Additional sources of information **are the employees themselves** and their personal experience.*

Another reflexive pronoun that causes difficulties for learners is *itself*. The analysis of concordance lines with *itself* showed that students tend to misuse it, for example:

(20) (a) *Therefore, the manager not only **itself** has to adhere to ethical standards, but also has to provide their observance in the organization in general.*

(b) *To develop a business model **itself**, I am going to use Osterwalder's techniques including Value Proposition Design, Prototyping and Business Processes Design.*

The reflexive pronoun *itself* can only refer to inanimate entities. In 18a it obviously refers to the antecedent *manager*, so the anaphor *himself/herself* should have been used instead. In Russian, to express the idea of sentence 20a the pronoun *sam* would be used, which has the same form for animate and inanimate entities. In contrast to English, Russian reflexive pronouns do not agree with their antecedent in number and gender, nor do they express animacy (Toldova et al., 2016). So the error, again, can be caused by the influence of the learners' native language. Secondly, the position of *itself* in the sentence does not appear to be quite normal. It would have been more natural to use it after the verb phrase *adhere to ethical standards*.

In 20b *itself* cannot be used after the noun phrase with the indefinite article. It would be correct to say *the business model itself*, but, having checked the context of use, it was concluded that the student did not mean a definite business model, so the reflexive pronoun in this sentence should be omitted.

The relative pronouns *which* was also sometimes used inaccurately in the learner corpus. For example:

(21) (a) *External marketing of the employer brand is the second step that is needed to attract potential employees, **which** may become loyal employees in the future.*

(b) *It is important to note, that some scholars, **which** investigations could be included at the present subgroup, consider employee displacement as a conception similar to the category of mass dismissals.*

In the author's view, errors in the use of relative pronouns can be caused by the influence of the students' native language. In Russian, the use of the relative pronoun *kotoryi* does not depend on the animacy of the antecedent.

The problem should be anticipated by teachers and the differences in the use of reflexive and relative pronouns in the two languages should be given thorough attention.

Verbal anaphora was also sometimes used incorrectly by the students. For example, in 22 to express the present tense meaning of the main verb, the anaphor *do* should be used.

(22) *It means that a company must have not only finished goods stocks like distributors **have**, but stocks of raw materials and spare parts, too.*

This might not be the influence of Russian but rather the insufficient attention verbal anaphora is given in ESL textbooks, as mentioned above.

Therefore, the reason for avoiding anaphora is likely to be the absence of specific teaching and the influence of the students' native language. Although some of the identified problems might appear quite basic and specific ESL issues (e.g. reflexives), the current research suggests the necessity of revising them in EAP courses, especially in relation to anaphora, in order to avoid confusion when later using anaphoric expressions in academic contexts. Special attention should be given to errors connected with the interference of learners' native language (e.g. the misuse of demonstratives and relative pronouns by Russian students). To tackle the problems, data-driven learning activities can be employed. Students can be given corpus examples from both experts' texts and their peers' writings. The former could demonstrate how anaphoric expressions serve to link parts of a text, as well as provide them with some usage patterns. The latter may be used for error correction and editing tasks.

Conclusion

This paper focused on referential coherence realised in academic texts with anaphoric expressions. Anaphora plays an important role in a text because it creates the cohesive chains (Halliday and Hasan, 1976) that are essential for text development. The author compared the use of anaphoric expressions in two corpora – a learner corpus that comprised L2 management students' writings and a reference corpus that included articles from peer-reviewed journals in management. The quantitative analysis provided evidence that there were significant differences in the use of anaphoric expressions in the two corpora, which confirmed the hypothesis. The qualitative analysis of the data identified four groups of problems learners had when using anaphoric expressions, some of which are in line with previous research (Leńko-Szimańska, 2004). The findings can be useful for academic writing instructors in terms of directing them to topics that deserve closer attention in EAP classrooms, as well as suggesting possible ways of dealing with the issues.

The author did not study referential coherence in other types of student writing, for example, academic essays, which is an obvious limitation of this work. Additionally, to make the task manageable, only a limited number of anaphoric expressions were examined. Another possible limitation of this research is that the observed differences in the use of anaphoric expressions may be due to the different contexts that the learner and expert texts were written in. Thus, the learners wrote their texts in a pedagogical context while the professional authors were knowledge producers. However, despite these limitations, the results from this study can shed light on some peculiarities of the way L2 students use anaphoric expressions in their academic texts.

The author believes that this paper is the first comparative study of learner and reference corpora focusing on referential coherence, and as such it may serve as a baseline for future studies. More data may be collected to extend both the learner corpus and the reference corpus. The collected data may also be compared to the data from a corpus of general English to analyse the use of anaphora in different types of discourse.

Acknowledgments

References

Al Sharoufi, H. (2014). A novel framework for teaching academic writing. *Pragmatics and Society, 5*(3), 484-507.

Basturkmen, H., & von Randow, J. (2014). Guiding the reader (or not) to re-create coherence: Observations on postgraduate student writing in an academic argumentative writing task. *Journal of English for Academic Purposes, 16*, 14-22. https://doi.org/10.1016/j.jeap.2014.07.005

Britton, B. K. & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology, 83,* 329-345.

Bunton, D. (1999). The use of higher level metatext in Ph.D theses. *English for Specific Purposes, 18,* 41-56.

Charaudeau, P., & Maingueneau, D. (2000). *Dictionnaire D'analyse du Discours* [Dictionary of Discourse Analysis]. Paris, France: Seuil.

Clark, H. H., & Haviland, S. E. (1977). Comprehension and the given-new contract. In R. O. Freedle (Ed.), *Discourse production and comprehension* (pp. pp. 1-40). Norwood, NT: Ablex.

Cornish, F. (2009). Inter-sentential anaphora and coherence relations in discourse: A perfect match. *Language Sciences, 31*(5), 572-592.

Crompton, P. (2017). Complex anaphora with this: Variation between three written argumentative genres. *Corpora, 12*(1), 115-148.

de Beaugrande, R. (1980). *Text, discourse and process: Toward a multidisciplinary science of texts*. Norwood, NT: Ablex Publishing Corporation.

de Beaugrande, R., & Dressier, W. (1981). *Introduction to Text Linguistics.* New York, NY: Longman.

Degand, L., Lefèvre, N., & Bestgen, Y. (1999). The impact of connectives and anaphoric expressions on expository discourse comprehension. *Document Design, 1*(1), 39-51.

Ducrot, O., & Schaeffer, J. M. (1995). *Nouveau Dictionnaire Encyclopŭdique des Sciences du Langage* [New Encyclopedic Dictionary of Language Sciences]. Paris, France: Seuil.

Dunning, T. (1993). Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics, 19*(1), 61-74.

Fahnestock, J. (1983). Semantic and lexical coherence. *College Composition and Communication, 34*(4), 400-416.

Ferris, D. R. (1994). Lexical and syntactic features of ESL writing by students at different levels of L2 proficiency. *Tesol Quarterly, 28*(2), 414-420.

Gabrielatos, C., & Marchi, A. (2012, September). Keyness: Appropriate metrics and practical issues. Paper presented at CADS International Conference 2012: Corpus-assisted discourse studies: *More than the sum of discourse analysis and computing?* (University of Bologna, Italy).

Gardelle, L. (2012). 'Anaphora', 'anaphor' and 'antecedent' in nominal anaphora: Definitions and theoretical implications. *Cercles: Revue Pluridisciplinaire du Monde Anglophone, 22*, 25-40.

Gardezi, A. A. & Nesi, H. (2009). Variation in the writing of economics of students in Britain and Pakistan: The case of conjunctive ties. In M. Charles, D. Pecorari & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 236-250). London, UK: Continuum.

Gilquin, G. (2000/2001). The integrated contrastive model. Spicing up your data. *Languages in Contrast, 3*(1), 95–123.

Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository Text? In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82– 98). New York, NY: Guilford Publications.

Granger, S. (1996). From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg & M. Johansson (Eds.), *Languages in contrast. Text-based cross-linguistic studies. Lund Studies in English* (vol. 88, pp. 37–51). Lund, Sweden: Lund University Press.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research,1*(1), 7-24. https://doi.org/10.1075/ijlcr.1.1.01gra

Grosz, B. J. & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics, 12*(3), 175-203.

Halliday, M. A., & Hasan, R. (1976). *Cohesion in English.* London, UK: Longman.

Halliday, M. A., & Matthiessen, C. (2013). Halliday's Introduction to Functional Grammar (4th ed.). Milton Park, UK: Routledge.

Hiltunen, T. (2016). Passives in academic writing: Comparing research articles and student essays across four disciplines. In M. López-Couso, B. Méndez-Naya, P. Núñez-Pertejo & I. M. Palacios-Martínez (Eds.), *Corpus Linguistics on the move: Exploring and understanding English through corpora* (pp. 132-157). Leiden, Boston: Brill.

Hoek, J., Zufferey, S., Evers-Vermeul, J. & Sanders, T. J. (2017). Cognitive complexity and the linguistic marking of coherence relations: A parallel corpus study. *Journal of Pragmatics*, *121*, 113-131.

Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing, 13,*133-151. https://doi.org/10.1016/j.jslw.2004.02.001

Hyland, K. & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics, 25*(2),156-177. https://doi.org/10.1093/applin/25.2.156

Kleijn, S., Pander Maat, H. L., & Sanders, T. J. (2019). Comprehension effects of connectives across texts, readers, and coherence relations. *Discourse Processes*, *56*(5-6), 447-464. https://doi.org/10.1080/016385 3X.2019.1605257

Lee, D., & Chen, S. (2009). Making a bigger deal of the smaller words: Function words and other key items in research writing by Chinese learners. *Journal of Second Language Writing, 18,*149–165. https://doi.org/10.1017/ S0267190512000037

Leńko-Szimańska, A. (2004). Demonstratives as anaphora markers in advanced learners' English. In G. Aston, S. Bemardini & D. Stewart (Eds.), *Corpora and language learners* (pp. 89-107). Amsterdam, Netherlands: John Benjamns.

Lesgold, A. M., Roth, S. F., & Curtis, M. E. (1979). Foregrounding effects in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior, 18,* 291-308.

Li, T., & Wharton, S. (2012). Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes, 11*(4), 345-356.

Martin, J. (1992). *English Text: System and Structure.* Amsterdam, Netherlands: John Benjamins. https://doi.org/10.1016/j.jeap.2012.07.004

Matysiak, I. (2007). Information extraction systems and nominal anaphora analysis needs. In *Proceedings of the International Multiconference on Computer Science and Information Technology (*pp. 183-192). Wisła, Poland: Polskie Towarzystwo Informatyczne.

Pavlou, P., & Potter, T. (1994). The difficulty of avoiding gender-biased language in highly inflected languages: A comparison of Greek and Arabic. Retrieved from https://files.eric.ed.gov/fulltext/ED371625.pdf

Plakans, L., & Gebril, A. (2017). Exploring the relationship of organization and connection with scores in integrated writing assessment. *Assessing Writing, 31*, 98-112. https://doi.org/10.1016/j.asw.2015.07.004

Rahimova, D. I., & Yusupova, Z. F. (2015). Semantics and pragmatics of demonstrative pronouns in Russian and Turkic languages. *Journal o fLanguage and Literature*, *6*(2), 113-116. https://doi.org/10.7813/jll.2015/6-2/22

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora* (vol. 9, pp. 1-6). Hong Kong, China: Association for Computational Linguistics.

Sanders, T., Land, J. , & Mulder, G. (2007). Linguistics markers of coherence improve text comprehension in functional contets. *InformationDesignJournal*, *15*(3), 219-235. https://doi.org/10.1075/idj.15.3.04san

Sanders, T. J., & Noordman, L. G. (2000). The role of coherence relations and their linguistic markers in text processing. *Discourse Processes*, *29*(1), 37-60. https://doi.org/10.1207/S15326950dp2901_3

Shaw, P. (2009). Linking adverbials in student and professional writing in literary studies: What makes writing mature. In M. Charles, D. Pecorari, & S. Hunston (Eds.), *Academic writing: At the interface of corpus and discourse* (pp. 215-235). London, UK: Continuum.

Smith, N. (2009). *Corpus Linguistics: A practical web-based course.* Available at https://www.lancaster.ac.uk/fss/ courses/ling/corpus/blue/l08_4.htm

Swales, J. M. (1990).*Genre analysis: English in academic and research settings.* Cambridge, UK: Cambridge

University Press.

Tanskanen, S. K. (2006). *Collaborating towards Coherence: Lexical cohesion in English discourse.* Amsterdam, Netherlands: John Benjamins Publishing.

Toldova, S., Azerkovich, I., Ladygina, A., Roitberg, A., & Vasilyeva, M. (2016). Error analysis for anaphora resolution in Russian: New challenging issues for anaphora resolution task in a morphologically rich language. In *Proceedings of the Workshop on Coreference Resolution Beyond Onto Notes (CORBON 2016)* (pp. 74-83). California, CA: Association for Computational Linguistics.

Vonk, W., Hustinx, L. G., & Simons, W. H. (1992). The use of referential expressions in structuring discourse. *Language and Cognitive Processes, 7*(3-4), 301-333.

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing ability. *College Composition and Communication, 32*(2), 189-204.

Appendix 1

Examples of exercises focusing on the identified problems with the use of anaphoric expressions in L2 writing.

Exercise 1. *Improve the sentences below using appropriate pronouns:*

1. For this reason, the research question is what factors affect the settlement period of troubled assets? In order to answer the research question, a series of tasks should be fulfilled.
2. Let us consider a step-by-step procedure of event study. The usual algorithm of the event study methodology consists of several steps.
3. Undoubtedly, it is impossible to use all existing methods in my study. Thus, the limitation of my study is the narrowed set of methods.
4. The second part of this work is devoted to advantages and disadvantages of one person companies which is subdivided into the comparison of one person companies in the Russian Federation and other countries, advantages of such companies and disadvantages of one person companies, the main problems created by the use of such companies.
5. Probably not all of the factors tested will appear to be significant factors.

Exercise 2. *Read the extracts from journal papers in management. How are ideas connected in the texts?*

(1) In a vignette experiment and follow-up interviews, we "introduced" a coercive demand to adopt a new practice in Dutch childcare under conditions in which employees and parent– customers either did or did not support the practice. We asked childcare managers when they would comply with the coercive demand (Study 1) and how they would respond to pressures from employees and parent– customers prior to compliance (Study 2).
(2) When an institutional pressure features force and coercion, the target of the pressure would appear to have little choice but to comply — and quickly. Yet, if other important constituents view the compliance behavior as illegitimate, the target is caught between conflicting constraints, or a proverbial rock and a hard place.
(3) Thus, need satisfaction reflects an organization's demonstration of its trustworthiness which therefore engenders employee trust in the organization. The satisfaction of needs should not only reduce uncertainty about the trustor's intentions or behaviours but also signal the trustor's investment in a social exchange relationship leading to trust in organization.

Exercise 3. *Correct the sentences below:*

1) The study of pricing strategies is an important practical issue, because it helps to understand how and why consumers are willing to pay for certain apartments, location factors and its characteristics.
2) It means, that some relations between parties may not be regulated by the law because this norms just don`t exist.
3) All companies should take proper care of its business-processes.
4) That approaches consider firm activities from a different perspective seeing it as complements, not substitutes.
5) There are five stages in this concepts: identification of value, identification of stream, alignment of stream, pulling and continuous improvement.

Appendix 2

List of the articles from the reference corpus cited in the paper.

AM 2015-1 Raaijmakers, A. G., Vermeulen, P. A., Meeus, M. T., & Zietsma, C. (2015). I need time! Exploring pathways to compliance under institutional complexity. *Academy of Management Journal*, *58*(1), 85-110.

AM 2015-3 Barrick, M. R., Thurgood, G. R., Smith, T. A., & Courtright, S. H. (2015). Collective organizational engagement: Linking motivational antecedents, strategic implementation, and firm performance. *Academy of Management journal*, *58*(1), 111-135.

AM 2015-6 Compagni, A., Mele, V., & Ravasi, D. (2015). How early implementations influence later adoptions of innovation: Social positioning and skill reproduction in the diffusion of robotic surgery. *Academy of Management Journal*, *58*(1), 242-278.

JMS 2014-8 O'Brien, E. and Linehan, C. (2014). A balancing act: Emotional challenges in the HR role. Journal of Management Studies, 51(8), 1257–85.

JMS 2016-8 Dai, Y., Roundy, P. T., Chok, J. I., Ding, F., & Byun, G. (2016). 'Who Knows What?'in New Venture Teams: Transactive Memory Systems as a Micro-Foundation of Entrepreneurial Orientation. *Journal of Management Studies*, *53*(8), 1320-1347.

Tables

Table 1

Frequency of nominal anaphora in the learner corpus and reference corpus

Nº	Referential marker	Absolute and normalized frequency (per 1000 words)		Log-Likelihood	%DIFF
		Learner Corpus	**Reference Corpus**		
		Subject pronouns			
	he	52 (0.400)	280 (0.403)	0.00	-0.86
	she	11 (0.084)	113 (0.162)	5.14	-48.03**
	it	459 (3.531)	1597 (2.301)	60.19	53.43***
	they	208 (1.600)	2453 (3.534)	150.90	-54.73***
	Total	730 (5.615)	4443 (6.402)	11.12	-12.29***
		Possessive adjectives			
	his	69 (0.530)	300 (0.432)	2.26	22.78
	her	17 (0.131)	156 (0.225)	5.20	-42.82**
	its	328 (2.523)	896 (1.291)	96.22	95.43***
	their	324 (1.800)	3788 (5.458)	229.12	-54.34***
	Total	738 (5.676)	5140 (7.406)	48.75	-23.35***
		Object pronouns			
	him	5 (0.038)	57 (0.082)	3.28	-53.17
	her	2 (0.015)	33 (0.048)	3.39	-67.65
	it	122 (0.938)	555 (0.800)	2.47	17.35
	them	122 (0.938)	653 (0.940)	0.00	-0.26
	Total	251 (1.930)	1298 (1.870)	0.21	3.23
		Reflexive pronouns			
	himself	1 (0.008)	14 (0.020)	1.15	-61.87
	herself	0	14 (0.020)	4.81	-100**
	itself	33 (0.254)	84 (0.121)	11.52	109.73***
	themselves	12 (0.092)	224 (0.323)	26.37	-71.40***
	Total	46 (0.353)	336 (0.484)	4.30	-26.91**

	Demonstratives			
this	1054 (8.107)	3768 (5.429)	122.50	49.33***
these	239 (1.838)	1996 (2.876)	48.05	-36.08***
those	67 (0.515)	698 (1.006)	32.87	-48.76***
Total	1360 (10.461)	6462 (9.311)	14.86	12.35***
	Indefinite pronouns			
one	20 (0.154)	30 (0.043)	16.87	255.90***
ones	7 (0.054)	55 (0.079)	1.02	-32.06
Total	27 (0.207)	85 (0.122)	5.19	69.57**
	Relative pronouns			
which	588 (4.523)	2252 (3.245)	48.11	39.39***
who	109 (0.838)	965 (1.390)	28.66	-39.70***
whose	13 (0.100)	89 (0.128)	0.75	-22.02
Total	710 (5.461)	3306 (4.763)	10.60	14.65***

Note: *** p<0.01, ** p<0.05 — significant at the 1% and 5% levels correspondingly.

Table 2
Frequency of verbal anaphora in the learner corpus and reference corpus

№	Referential marker	Absolute and normalized frequency (per 1000 words)		Log-Likelihood	%DIFF
		Learner Corpus	Reference Corpus		
	do	5 (0.038)	151 (0.218)	26.08	-82.32***
	does	1 (0.008)	8 (0.012)	0.16	-33.27
	did	2 (0.015)	29 (0.042)	2.51	-63.18
	Total	8 (0.061)	188 (0.270)	27.26	-77.82***

Note: *** p<0.01, ** p<0.05 — significant at the 1% and 5% levels correspondingly.

Table 3
Frequency of adverbial anaphora in the learner corpus and reference corpus

№	Referential marker	Absolute and normalized frequency (per 1000 words)		Log-Likelihood	%DIFF
		Learner Corpus	Reference Corpus		
	here	15 (0.115	194 (0.280)	14.09	-58.72***
	there	32 (0.246)	174 (0.251)	0.01	-1.82
	Total	47 (0.361)	368 (0.530)	6.75	-31.82***

Note: *** p<0.01, ** p<0.05 — significant at the 1% and 5% levels correspondingly.

Table 4
The most common patterns with themselves in the reference corpus

Pattern	Absolute Frequency	Example
see / saw themselves as + a noun / adjective / present participle	8	The second group also saw themselves as making a sacrifice (JMS2014-51)
a verb + themselves and others	8	With this added perspective, scholars might have a better grasp on how people perceive themselves and others in the context of a status change (AM 2016-59).
to make themselves + a verb / more + an adjective	5	Some scholars have noted that firms engage in downsizing to make themselves more attractive to potential buyers (JM 2013-39)