TEACHING NOUN-PHRASE COMPOSITION IN EAP/ESP CONTEXT: A CORPUS-ASSISTED APPROACH TO OVERCOME A DIDACTIC GAP

Maria Malakhovskaya¹, Larisa Beliaeva¹, Olga Kamshilova¹,²
¹Herzen State Pedagogical University of Russia, Russia
²Saint-Petersburg University of Management Technologies and Economics, Russia

Abstract. Attributive noun phrases, namely noun + noun constructions (NNCs), is an essential typological characteristic of academic style. However, errors in their production and use are not uncommon even in the writing of advanced Russian users of academic English. As such errors slow down reading and, in some cases, lead to misinterpretations, they may be a hindrance to the publication of research results in international journals. Aiming at facilitating NNC instruction, the study focuses on linguistic and didactic aspects of this construction. It highlights NNC complexity and a lack of consensus among researchers as to their linguistic status and semantic nature. The research demonstrates that NNCs do not find proper interpretation in EAP/ESP teaching literature, which results in their misuse by non-native English contributors to international scientific journals. Based on some published practice and personal experience in teaching academic skills, the research lays down a corpus-assisted approach to teaching students to master the grammar and use of NNCs in their own discipline texts. The principal advantage of this approach is that it provides students with patterns of NNP expert use and aids EAP/ESP instructors to develop discipline-specific didactic materials based on reliable and up-to-date facts on generating NNCs.

Key words: noun phrase (NP), noun+noun construction (NNC), corpus technologies, syntactic complexity, English for Academic purposes (EAP), English for Specific Purposes (ESP)

1. INTRODUCTION

Ever since the role of English as a lingua franca was generally recognized in international scientific and academic communication there has been a boom in the field of educational literature on academic English (EAP) and English for specific purposes (ESP) studies. This inspired a growing research interest in the processes of teaching academic writing and academic literacy skills. Academic literacy skills involve working on two “parallel but related levels”, which were defined as discursive literacy and linguistic literacy (Liu and Li 2016, 49). Discursive literacy is the ability to create a scientific text in accordance with the logical, structural and stylistic requirements that...
have developed in each discipline for texts of various genres and types. Linguistic literacy implies adequate lexical and grammatical skills. Teaching discursive literacy is axiomatic, especially for academic and scientific communities whose national scholarly writing traditions differ from today’s international discursive practices. However, some research findings indicate that non-native English (L2) writers of scholarly texts see their problems rather at the linguistic level (i.e. they are concerned about a lack of skills in using proper vocabulary and syntax to express their ideas), than at the global, discursive level (Qian and Krugly-Smolska 2008; Liu and Li 2016). These findings cannot be neglected and the linguistic component of an academic course curriculum should include discipline-specific lexical and grammatical constructions.

One such construction is a multi-word noun phrase (multi-word NP), a syntactic construction with one or more nominal premodifiers, that is $N_1+N_2+...+N_n$ constructions (NNCs hereinafter). As we will later show, this construction, common as it is in scientific text, has not been properly addressed in academic English courses. As a result, even texts by advanced L2 users of English are not deprived of errors in multi-word noun phrase composition. Such errors can significantly affect the perception of their texts by readers, including reviewers of international academic journals.

The focus of the paper is on complexity of NNC syntax which disables ESP/EAP learners and writers for international journals both to understand and compose multi-word NNCs. We claim that NNCs are text-dependent and discipline dependent units. Hence, one way to deal with them in EAP / ESP instruction is to develop discipline-specific teaching techniques based on the use of linguistic corpora. This study will outline some approaches to developing such methodology.

2. NNCs in Scientific Texts: Syntactic, Didactic and Contextual Aspects

2.1. NNCs as a marker of scientific text syntactic complexity

The well-known classical work of Biber et al. (2007) states that 30% of all noun premodification structures in academic texts are NNCs, their presence in texts of other register/style being considerably lower (ibid., p.589). It means that this construction is an essential typological characteristic of scientific style. Therefore, it should be properly represented in the EAP / ESP curricula. However, due to the complexity of the semantic and syntactic nature of NNC, the development of teaching materials on this topic would not be possible without a careful consideration of the didactic and linguistic aspects of this phenomenon.

A large number of recently published studies concentrate on didactic aspects of NNCs. Most of them follow Biber, Gray and Poonpon (2011), who assumed that the level of academic literacy correlates with the choice of language tools that students choose to express attributive semantics: students of a beginner level would most frequently use adjectives and relative clauses as attributes, and later, as they improve in academic writing, they start using more complicated syntactic structures based on multi-word NPs.

The hypothesis was later proved by numerous researchers, who showed that texts of authors who are more advanced in academic writing, regardless of whether English is their first or foreign language, are characterized by a much denser use of complex NPs than those of less experienced authors (Parkinson and Musgrave 2014; Ansarifar et al. 2018).
It is well known that syntactic complexity is one of the requirements for academic writing (Ansarifar et al. 2018). It is perceived as an indicator of the writer’s belonging to the academic community both by readers and peer reviewers. That is why teaching students how to create syntactically complex texts is a most important task facing teachers of academic writing.

NNCs play a special role in writing syntactically complex scholarly texts for two reasons. Firstly, the use of nominal premodifiers is a common way to produce new terminological units (e.g., climate change, sex education, identity formation, assertion component, responsive web design). Secondly, the implicit semantic structure of such terms is not always transparent and needs to be explicated through the context in which NNCs are used in each subject area. Hence, good skills in NNC use will indicate a student’s progress in the subject area (Elliott, 2019).

Thus, it is both the frequency of NNC in academic text and its syntactic complexity, that leads us to a conclusion that this structure should be seriously considered in EAP/ESP curricula. It is all the more necessary as these syntactic complexes present problems in instruction due to their controversial grammatical status and complicated semantic nature. Our teaching experience gives much evidence of how challenging NNC grammar is for Russian ESP/EAP learners and how short of relevant didactic support the teachers are.

2.2. NNCs in ESP/EAP instruction

Although much research has been done on the patterns of NNCs acquisition by non-anglophone academic writers, so far very few studies have highlighted the errors L2 authors tend to make in composition and use of these constructions and the problems teachers face when teaching them (Peter Strutt’s Delta Pages 2015). We analyzed various types of NNC misuse in academic texts, written by researchers and postgraduates of Herzen University (Saint Petersburg, Russia). The authors’ pretested language proficiency was B2–C1. The analyzed texts were records of English lecture courses and draft versions of articles intended for publication in international scientific journals. The analysis revealed the authors’ awareness of NNC role in a scientific text, though composition of NNCs in their texts was very often incorrect, unnatural, and, as a result, incomprehensible for the reader. Here are most frequent mistakes:

1) Globalization affects culture life in our country вместо (instead of cultural life);
2) Now we have to consider the countries differences (instead of difference between the countries);
3) Children literature is of great importance in preschool education (instead of children’s literature).
4) (in the title) Three design theory (instead of The theory of three designs).

The illustrations show, that Russian ESP/EAP learners tend to overuse NCCs, prefer a noun component as a premodifier even if there is a suitable adjective, and use plural nouns as premodifiers. Similar tendencies are found in academic texts produced by speakers of other languages (Peter Strutt’s Delta Pages 2015).

One may assume, that to minimize such mistakes a teacher should support the students with a clear and coherent system of different forms expressing attributive relations, including both prepositive structures (NNCs and possessive nouns) and postpositive attributes with prepositions. Such a system might help them understand their functional and context
differences and, in particular, indicate restrictions on NNC compositions and use. But as it turns out, the teacher has almost nothing to rely on for two reasons which we will address in the next sections.

2.2.1. NNC and other noun modification constructions in coursebooks and grammar reference books

A survey of most popular courses on teaching English for general purposes revealed that some of them do not focus on grammar of attributive relations at all (New Cutting Edge, New Inside Out), others would give most general commentaries, limited to a page or two (Straightforward, English File). A more detailed exposition of this topic can be found in advanced grammar books and grammar reference books (Oxford English Grammar in Use Advanced, Advanced Grammar in Use by Martin Hewings). But even there no explanation is given of NNC place among other means of expressing attributive relations and no guide for their use or possible limiting rules. Besides, the prescriptions given there either differ from real language use, or cannot be followed because of students’ lack of language and cultural experience. A prescription of that type is found, for instance in (Hewings 2008, 86). It recommends using NNCs only for naming “a well-known class of items”. For instance, according to the book, a compound nominal construction income tax is quite natural, while a combination like *children’s clothes tax is not and needs turning into a postmodifying prepositional construction tax on children’s clothes. This lapidary comment undoubtedly does not give much help and only makes the rule on NNC use less clear. Even if NNCs were nominations of well-known classes of items alone, the prescription would be of little use for ESP/EAP learners, since, being speakers of other languages, they could hardly know which classes of objects are well-known and shape integral concepts in native speakers’ world picture.

Coursebooks on academic English do not give this topic due attention, either. For instance, Oxford grammar for EAP considers only a very limited list of semantic relations between NNC components and there is no comparison of NNC with other ways of expressing attributive relations. They advise to use it widely but there is no comment on its possible misuse: “It is a technique you should consider using wherever you can” (Paterson and Wedge 2013, 32).

The fact, that a grammar topic so important for academic writing is scarcely mentioned even in specialized coursebooks, explains the numerous mistakes students make in NNC production. The NNC grammar is not taught to students, whatever proficiency level is concerned. Means of expressing attributive relations are fragmentary, not systemic, being neither compared, nor contrasted to each other (Peter Strutt’s Delta Pages 2015). The relation types of NNC components are rather superficially analyzed. As a result, a learner is convinced that they can produce NNCs by mere placing nouns in a line.

2.2.2. NNC grammatical status and semantic interpretation as a challenge in ESP/EAP instruction

These gaps in ESP/EAP curriculum may be explained by the fact that NNC is a highly complex language phenomenon, and even today its linguistic description demonstrates neither a complete understanding, nor a consensus of opinions.

Firstly, NNC grammatical status remains undetermined. Many linguists argue that it is impossible to strictly differentiate two-component nominal structures from free word
groups. They state that NNCs are on the cusp between morphology and syntax (Bauer 1998; Fernández-Domínguez 2010). Hence, NNC terminology varies from nominal compounds to noun-noun combinations, or noun+noun sequences, etc. (see Linh 2010). This may objectively reflect problems in NNC linguistic interpretation, but in terms of teaching, it is nothing but a handicap, a hindrance that leads to unconvincing recommendations of the kind we discussed above.

Semantic relations between NNC components is another point of doubt and controversy. Lists of defined semantic relations between NNC components differ both in number and types. Thus, Biber speaks of 15 types of semantic relations between two nouns in a 2-component NNC – composition (leather coat), purpose (war fund), source (whale meat), time (summer holiday), location (heart disease), etc. (Biber et al. 2007). Other linguists present nomenclatures either shorter or longer than that (see Fernández-Domínguez 2010), and still others claim that such classifications have no sense at all (see Linh 2010). Teaching NNC grammar might be more effective if a teacher had a reliable classification of semantic relations within NNC structure. If they had one, it would be possible to use it as a guide for NNC production and advise students to discard the constructions that do not fit the classification patterns.

One more difficulty in NNC learning and teaching is the lack of morphological elements that could mark the syntactic dependency between the head word and the modifiers. The NNC external simplicity is misleading and most often an NNC meaning can be defined only by its context (Fernández-Domínguez 2010). NNC semantic ambiguity is often illustrated by a woman doctor (a doctor for women or a doctor who is a woman).

So, teaching NNC to ESP/EAP students must become an issue of serious didactic concern. Since acquiring skills in its use and composition is such a challenging task due to its informational density, variety of semantic and syntactic relations between its components and its potential ambiguity (Elliott 2019), it is crucial that it should be a significant component of ESP/EAP curriculum.

2.2.3. NNC variation across various subject areas

In recent years, a number of studies have been published that reveal significant discrepancies in the use of language tools in texts belonging to different disciplines, genres, and types. Some researchers argue that language characteristics of academic texts depend, to a greater extent, on the subject area or discipline, than on the author’s native language and cultural background (Adel and Romer 2012). NNCs have also become objects of comparison across different subject areas.

Elliott (2019), for instance, analyzed the NNC use in The Michigan Corpus of Upper-Level Student Papers (MICUSP). He came to a conclusion that NNC frequency is in direct dependence on the discipline the text belongs to. The highest frequency was found in texts on hard disciplines, while the lowest is characteristic of soft disciplines. He also demonstrated that NNC frequency depends on whether the discipline is pure or applied. Applied disciplines show more NNCs than pure ones. Besides, he found that the length of noun strings is also discipline dependent: the soft-pure disciplines show significantly fewer noun strings with two or more nouns as premodifiers than the hard-applied ones.

On the whole, research like this highlights the value of a discipline-specific approach to teaching noun+noun sequences to ESP/EAP students. We suggest that such approach can be provided with corpus technologies, which may be used in class or individually, on the one hand, and as a resource for new teaching materials, on the other.
3. CORPUS TECHNOLOGIES IN NNC STUDY AND PRACTICE

The rise of corpus technologies provides new possibilities both for research and education. The use of text corpora and corpus tools has proved effective enough in helping the learner to enrich their vocabulary and master the grammar and style of their own discipline texts. Basing on some published practice (Linh 2010; Anthony 2019) and personal experience we attempt to describe some corpus-assisted practices that may be applied to teach students to deal with NNCs in texts of their subject area.

Following our claim that an NNC is rather a unit of syntax, not lexicon, and that its syntactic organization is much related to the context, we suggest to overcome learners’ difficulties with NNC use and production by referring to text corpora, both well-known and original, built by the learner and including texts of their own discipline.

3.1. Method

The methodology is based on the use of computer tools for text analysis, namely AntConc, which is a free and reliable corpus manager, a toolkit with a number of useful instruments for text processing. The choice of the tool is determined by a number of reasons among which are the tool availability and reliability, and a relatively small volume of the processed corpus. Corpus processing mainly involved such instruments as Word List, Concordance, Clusters/N-Grams, and the procedure included the following steps:

Step 1. Collecting and normalizing texts in learner’s own discipline: titles and non-text elements (figures, pictures, hyperlinks, etc.) are to be deleted, txt-format is recommended to load the text in the instrument;

Step 2. Generating the corpus Word List. Scrolling down and exploring the data (after the stopwords are deleted) finds most frequent meaningful words, that can be sorted by Frequency, Word (alphabetical order) or Word end;

Step 3. Concordancing. Examining a word in its immediate context; the tool is useful for establishing the length of NNCs and the position of a word in an NNC (head or modifier)

Step 4. Extracting Clusters/N-Grams. The tool helps to extract bi-Grams, tri-Grams, etc. and determine their frequency in the text. It also demonstrates gradual growth of an NNC within a text.

Interpreting the corpus findings (N-Grams, clusters and concordances) helps to judge upon the role of text in interpreting syntactic relations within an NNC and its meaning.

3.2. Procedure: searching NNCs in a corpus

Collecting data. A learner of ESP/EAP may find helpful searching web-based corpora, such as RCPCE collection of profession specific corpora as well as national corpora and corpora embedded in online text analysis tools like Sketch Engine. For instance, RCPCE collection contains corpora from 39 disciplines which can be processed by a number of tools that allow search for individual words/phrases or two or more associated words/phrases and

---

1 Stopword lists and lemma lists can also be added to the tool, for example, PubMed's list of 132 stopwords. Many such lists exist for reuse on the internet and the choice depends on the context of the search - McGill Library – URL: https://www.mcgill.ca/library/

2 Research Centre for Professional Communication in English – URL: http://rcpce.engl.polyu.edu.hk/rcpce/index.html

3 Sketch Engine – URL: https://app.sketchengine.eu/
Teaching Noun-Phrase Composition in EAP/ESP Context

It is also possible to add texts that a learner can find on their own. Working with AntConc is exclusively aimed at their individual choice, that means the collection may be restricted to learners’ specific discipline and include articles written by experts in the field, which is more effective according to recent studies (see section 2.2.3). The illustrations below are based on searching a small corpus of text on terminological studies.

After the collection has undergone normalization stage it is added to the tool, previously set for proper character encoding, text by text. AntConc can read several text formats: .txt, .html, .xml, .ant. We recommend .txt file as most suitable.

Word List. Scrolling down the Word List helps to find most frequent nouns which, in most cases, are terms or term components describing the subject area. A high frequency noun can be actualized in the text either as one-component NP or as the head noun in a multicomponent one, that is why every noun with frequency 2 or more shall be fixed and analyzed in its immediate context. Let us consider segmentation as a candidate for an NP (NCC) component, as its frequency in a processed text on terminological issues is 6.

Clusters/N-Grams. This tool finds clusters with the search word, words immediately adjoining it in the text. The length of n-grams with the search word is set in the box “N-Gram Size” in accordance with research goal. The tool output for segmentation is 6 different tetragrams with segmentation. The noun is used in the text as one-component NP only once, in 4 other cases it is part of a premodified NNC: segmentation rules (3) and segmentation rules exchange (1). There is one more combination in the output – segmentation (tr# 29 grapheme boundaries) – to define the status of which, as well as of the other 4, the Concordance tool shall be applied.

Concordance. Concordance is a tool that shows how the NP/NNC is used in real text, demonstrating its immediate context. The context analysis confirms that segmentation is a noun-premodifier in segmentation rules, the latter premodifies exchange in a 3-component NNC segmentation rules exchange. To decide on the status of segmentation (tr# 29 grapheme boundaries) the concordance line is not enough, for that purpose it is necessary to use File View tool.

File View. Clicking File View box provides the researcher with a broad context, a text fragment that allows to see, as in the case in question, that one of the tetra-grams for segmentation cuts a chunk that is not a grammatical sequence at all, since there is a boundary marked by a bracket introducing a syntactically loose part: segmentation (tr# 29 grapheme boundaries).

Thus, a researcher or student may find actually and expertly used NPs of different composition and length in texts of their specific discipline. Corpus tools provide reliable facts on generating multi-component NPs and NNCs, give patterns of their expert use and, in that way, may help to compensate the deficiency of didactic materials that we discussed in 2.2.1.

NNC analysis within the text space leads to establishing procedures of coining novel NNCs from those featuring in the text and to recognizing the compressed sentence structure in a concise form of an NNC.

4. DISCUSSION

As it has been shown, NNCs present a challenge for L2 authors and require a careful consideration in EAP and ESP instruction. In this regard, corpus methodology might be a useful tool due to its potential to provide non-native English speakers with information
they need to identify and unpack these constructions in scientific texts as well as to use them in their own texts. However, when teaching this topic, it is necessary to emphasize the dual nature of this construction, i.e. the fact that its conciseness, which is so effective in scientific texts, is achieved at the expense of explicitness. The resulting ambiguity “places an increased burden on readers” (Elliott 2019, 12). This may not be much of a problem for the specialist audience, but given the current interest in multidisciplinary research, authors should consider the difficulties that less experienced readers are likely to face when trying to extract information from noun + noun premodification sequences. This consideration is especially important when it comes to supporting L2 authors in their efforts to prepare manuscripts for publication in international journals. As Wallwork noted, it is often the case that papers are reviewed by referees who are not top experts in the field, so the best strategy is to “write in a way that a non-expert or less experienced person can understand” (Wallwork 2013, 13-14). Obviously, texts overloaded with informationally-dense but ambiguous noun phrases cannot meet this requirement.

The potential ambiguity in noun premodification should be taken into account in teaching academic writing. Students need to be instructed how to achieve a balance between informational density and clarity. It means that they have to be aware of other ways of noun modification in academic English. Some researchers suggest that one of the sound strategies used by proficient academic writers to create dense information structure while establishing explicit relations between NP components is using complex post-modifying prepositional phrases (Biber and Gray 2016, 191). Not only does prepositional postmodification enable the writer to transmit information clearly and efficiently but it also contributes to the syntactic complexity of the text: some types of post-modifying prepositional phrases are associated with the highest developmental stages for noun phrase modification proposed by Biber et al. (2011). Therefore, it is important that L2 writers’ repertoire of noun modification be expanded by post-modifying prepositional phrases.

5. CONCLUSION

NNCs are an essential typological characteristic of academic style. They are the instruments of information density and markers of syntactic complexity, so the ability to produce correct NNCs indicates a high level of the writer's academic literacy. However, our analysis of NNC use in scholarly texts written by Russian researchers and postgraduates has revealed that even advanced users of English make numerous mistakes in coining NNCs, generally tending to overuse them. As a result, noun sequences in their texts are often incorrect, unnatural, and incomprehensible for the reader. This confirms the findings of some other researchers, which demonstrate that the use of NNCs is challenging for non-native English writers.

A critical survey of popular coursebooks and reference books has proved that they give no explanation of NNC place in the system of noun modification and no guide for their use or possible limiting rules. The study also has highlighted the complexity of the NNCs and a lack of consensus among researchers as to their linguistic status and semantic nature.

The fact that a grammar topic so important for academic writing is scarcely mentioned even in specialized coursebooks urges ESP/EAP teachers to develop their own educational materials. We suggest that one way to support learning about NNC composition and use is to apply corpus technologies with respect to NNC specificity in different subject areas.
Our methodology is based on the use of reliable computer tools for text analysis, such as AntConc. We have worked out a four-step procedure which can be adjusted to learners’/researchers’ individual needs. It means that the collection of data may be restricted to learners’ specific discipline and include papers written by experts in the field. Following this procedure, a researcher or student may find actually and expertly used NPs of different composition and length in texts of their specific discipline.

Corpus tools provide reliable facts on generating multi-component NPs and NNCs, give patterns of their expert use and, in this way, may help to compensate for the deficiency of teaching materials in NNC instruction. They also enable users to establish procedures of coining novel NNCs from those featuring in the analyzed text space.

This methodology may open a pathway to those concerned with the gap in ESP/EAP curricula as regards NNC instruction. The suggested procedure may enable them to create discipline-specific teaching materials and provide students with both well-structured and consistent information on the use of this construction and a system of relevant practical activities.

REFERENCES


