

## **CORPUS-BASED APPROACH IN LEARNING COLLOCATIONS IN INFORMATION-COMMUNICATION TECHNOLOGIES (ICT)**



**<https://doi.org/10.5281/zenodo.7670936>**

**Durdona Kadirbekova**

PhD in Philology, associate professor,  
Head of the “Foreign languages” department,  
Branch of the Russian State University Oil and Gas  
named after I.M.Gubkin in Tashkent city, Uzbekistan

E-mail: [d\\_kadirbekova@mail.ru](mailto:d_kadirbekova@mail.ru)

+99890 974 23 79

### **ABSTRACT**

*English for specific purposes (ESP) teachers often face dilemmas in deciding what lexical items to teach their students. In the field of ICT, there is no exception to this issue. The ICT corpus made up of research articles can provide better insights and guide to both ICT students and teachers. In EFL contexts, English for specific purposes often has great significance for learners, mainly due to its multi-disciplinary nature, and for advanced learners, collocations are especially important. As Özdemir (2014) notes, concordance lines allow us to observe potential specialized collocations. For example, collocations identified from ESP corpora have served to facilitate learning for disciplines such as nursing, medicine, and tourism; however, no corpus yet exists for Information-Communication Technologies (ICT). Creation of this ICT corpus directly assists instructors to identify collocations in the field, furthering communication, report writing, and interpretation.*

*The goal of the study is identify, analyze and compile a list of specialized collocations that occurred frequently in a 803,294-word ICT corpus. Homemade specialized corpus for ICT is presented. Implications for research and pedagogy are discussed. Ant Conc helps to learn new words that can be found in certain professions, the functions of the corpus also help to recognize the meanings of words and the difference between synonyms, so that students use the most semantically appropriate words in speech and construction of texts. The use of the corpus helps to diversify the vocabulary, replenish it with professional terms. We believe that the corpus will be implemented in the education process of Foreign Languages, which will regulate future specialists career growth and the development of professional competencies, which will undoubtedly be reflected in the future development of the ICT industry of the Republic of Uzbekistan.*

**Keywords:** *corpus building, ICT collocations, corpus-based vocabulary list, teaching English language, English for Specific Purposes, research journals in ICT.*

## **AXBOROT-KOMMUNIKATSION TEXNOLOGIYALARI (AKT)GA OID SO'Z BIRIKMALARNI KORPUS ASOSIDA O'QITISH**

**Durdona Kadirbekova**

filologiya fanlari bo'yicha PhD, dotsent,

“Chet tillar” kafedrası mudiri,

I.M.Gubkin nomidagi Rossiya davlat neft va gaz

universitetining Toshkent filiali, O'zbekiston

E-mail: [d\\_kadirbekova@mail.ru](mailto:d_kadirbekova@mail.ru)

### **ANNOTATSIYA**

*Maxsus maqsadlar uchun ingliz tili (ESP) o'qituvchilari o'z talabalariga qanday leksik birliklarni o'rgatish kerakligini hal qilishda ko'pincha dilemmalarga duch kelishadi. Axborot kommunikatsiya texnologiyalari (AKT) sohasida bu borada istisno yo'q. Tadqiqot maqolalaridan tashkil topgan AKT korpusi AKT talabalar va o'qituvchilari uchun yordamchi manba bo'lishi mumkin. EFL kontekstlarida, aniq maqsadlar uchun ingliz tili ko'pincha o'quvchilar uchun katta ahamiyatga ega, asosan uning ko'p fanli jihati tufayli o'quvchilar uchun so'z birikmalarni bilish ayniqsa muhimdir. O'zdemir ta'kidlaganidek, muvofiqlik chiziqlari (concordances) ixtisoslashgan maxsus birikmalarni kuzatish imkonini beradi. Misol uchun, ESP korpusidan aniqlangan birikmalar hamshiralik, tibbiyot va turizm kabi fanlarni o'rganishni osonlashtirishga xizmat qilgan; ammo Axborot-kommunikatsiya texnologiyalari uchun korpus hali mavjud emas. Ushbu AKT korpusini yaratish bevosita o'qituvchilarga sohaga oid hamkorlikni kengaytirish, aloqani rivojlantirish, hisobot yozish va talqin qilishda yordam beradi. Tadqiqotning maqsadi 803 294 so'zli AKT korpusida ko'p uchraydigan maxsus ixtisoslashtirilgan birikmalar ro'yxatini aniqlash, tahlil qilish va tuzishdir. Maqolada AKT uchun yaratilgan ixtisoslashtirilgan mahsus korpus bayon etiladi. Tadqiqot va pedagogika uchun ahamiyati muhokama qilinadi. Ant Conc ma'lum kasblarda uchraydigan yangi so'zlarni o'rganishga yordam beradi, korpusning funksiyalari ham so'zlarning ma'nolarini va sinonimlar orasidagi farqni aniqlashda yordam beradi, shuning uchun o'quvchilar nutqda va matnlarni yozishda semantik jihatdan eng mos so'zlardan foydalana oladilar. Korpusdan foydalanish lug'at yaratishda, mavjud lug'atlarni professional terminlar bilan to'ldirishga yordam beradi. Ushbu maqolada biz AKT tadqiqot maqolalarining ESP korpusi orqali AKTda tez-tez qo'llaniladigan birikmalarni aniqlash bilan bog'liq jarayonlarning tavsifini taqdim etdik. Chet*

*tillarini o'qitish jarayonida bo'lajak mutaxassislarning kasbiy kompetensiyalarini rivojlantirishni tartibga soluvchi muhim chora-tadbirlarni amalga oshirishda shu kabi korpus katta ahamiyatga ega va bu O'zbekiston Respublikasi AKT sohasining kelajakdagi rivojlanishida shubhasiz o'z ifodasini topadi.*

**Kalit so'zlar:** korpus qurish, AKT so'z birikmalari, korpusga asoslangan lug'atlar ro'yxati, ingliz tilini o'qitish, maxsus maqsadlar uchun ingliz tili, AKT bo'yicha tadqiqot jurnallari.

### **Acknowledgment**

*This research's has been presented in TESOL 2019 international convention and was discussed by specialists. Research model was recommended to investigate vocabulary scope of other disciplines. The feedbacks were taken account and research methods, data interpretation have been improved. We wish to thank Dr. Gena Bennet for her valuable comments on different drafts of this work. Dr. Bennet is person who inspired this research and provided relevant knowledge regarding ESP corpus building. This work was supported by project 'Scholarly Research and Publication for ELT in Uzbekistan' conducted by the USA Embassy in Tashkent and Republican Scientific - Practical Centre for Development Innovative Methods of Teaching Foreign Languages under the Uzbekistan State World Languages University.*

### **INTRODUCTION**

Integration into the world community, conditions of accelerated development and globalization of information and communication technologies have sharply increased the need to learn foreign languages. In this regard, the introduction of advanced approaches of learning vocabulary into the education system is particularly important in the training of specialists who are fluent in foreign languages and are sufficiently aware of the latest achievements of science and technology. A new turn in linguistics is the section of linguistics – corpus linguistics, the development of which is closely connected with the digital industry. There has been a sharp leap in the development of computer and information technologies. These possibilities began to be successfully used in linguistics. Thanks to the development and popularization of the global Internet, a huge number of users from different countries could use the data from the corpus [Kallas E., Koppel K. R., Kallas R. G., 2017; 195]. A linguistic, or language, corpus of texts is understood as "a large, electronically presented, unified, structured, labeled, philologically competent array of linguistic data, designed to solve specific linguistic problems" [Maiorova A. 2017; 42-46].

The teaching and learning of English for ICT purposes can be enhanced with the existence of the ICT corpus. The initial function of an ICT corpus should directly

assist specialists in understanding the language for the purpose of communication i.e., corpus linguists and ESP teachers will be able to compare the results taken from this corpus with textbooks and other materials; researchers from the seven sub-disciplines of ICT (telecommunication, information security, computer engineering, software engineering, TV technology, radio technology and mobile communication) will be able to be aware of the certain words frequently used in the recent scholarly research articles in their field of study; they will also be able to know how extensive a certain term or theme has been written in the recent scholarly research articles in their field of study; academic English language teachers will be able to know the common collocates of a certain word when it is used in scholarly research articles; translators will be able to choose the appropriate words and collocations when translating a certain phrase from one language into English.

In English as a foreign language contexts, English for specific purposes (ESP) often has great significance for learners, mainly due to its multi-disciplinary nature, and for advanced learners, collocations are especially important [Abedi Z. Mobaraki M., 2014; 632], and as Özdemir notes, concordance lines allow us to observe potential specialized collocations [2. Vol.3. P. 37]. For example, collocations identified from ESP corpora have served to facilitate learning for disciplines such as nursing, medicine, and tourism [Abedi Z. Mobaraki M., 2014; 633; Ozdemir N., 2014; 21] however, no corpus yet exists for Information-Communication Technologies (ICT). Creation of a specialized corpus directly assists instructors to identify collocations in the field, furthering communication, report writing, and interpretation [Mohamad A., 2013; 21]. It is important to note here that the identification of the most frequent field-specific academic collocations is essential [Chung T., 2003; 103] for responsible planning of ESP or English for Academic Purposes courses. The field of English for Academic Purposes such as English for ICT purposes has become vital in bridging the ICT students from their general English into learning the language in ICT field. This issue has become more significant when students enroll to higher education system in Uzbekistan. Most of the ICT institutions in this country have developed an intensive program with the hope to solve this matter. Unfortunately, the English instructors who are responsible in teaching ICT students may also encounter difficulties; not knowing the vocabulary to be focused on as they are not from the ICT background. As number of scholars stated, having an appropriate knowledge of academic vocabulary has been accentuated the success at high level of education. [Coxhead A., 2000; 213; Nation I., 2001; 6].

Furthermore, corpus based learning of linguistics can be defined as “...the study of a language based on examples of ‘real life’ language use” [Nation P., Waring R., 1997; 15].

The best way to find out this matter would be by looking at a corpus of the field in order to define the way the collocations are presented in the ICT research scope text. On the other hand, the ICT field is exclusive in the sense that they have a specialized set of collocation which differs with its distinctions from other fields [McEnery T., Wilson a., 2001; 80]. Accessibility of collocations can be helpful to solve some problems related to the production of correct lexical units. It is obvious that there are various points of views on defining, collecting, encoding, organizing printed or electronic reference works [L’Homme M., 2009; 2]. The analysis of the academic English language corpus is able to characterize the usage of vocabulary, their frequency and range coverage of the texts in different fields which lead to the formation of word lists [Nation I., 2001; 61]. This work deals with the collocations used in an ICT scholarly research articles of 803294 tokens (running words) were found from 140 articles in 23 scientific journals. The journals were classified into seven sub-disciplines of ICT: telecommunication, information security, computer engineering, software engineering, TV technology, radio technology and mobile communication. The pre-determined criteria for the selection are: 1) The journals do not appear in more than one subject area; 2) the journals have a 5-year impact factor; 3) the articles are published in 2015 and 2018; 4) the articles are written in English; 5) the articles are open access.

### 1.1. Value of Specialized Vocabulary in ESP

In the world, research is being conducted in a number of priority areas, including defining specialized vocabulary; improving innovative technologies based on choosing vocabulary for English classes; developing approaches to select appropriate vocabulary list for ESP aimed at ensuring consistency of national and international best practices; improving technologies to eliminate difficulties in the perception of a heavy load of corresponding specialized vocabulary.

The English language is a language for international information interchange in every sphere. So, the ESP approach has been distinguished from general English in language teaching [Hutchinson T., Waters A., 1987; 12]. A big amount of specialized vocabulary or “... technical words that are recognizably specific to a particular topic, field, or discipline” is one of the unique features of ESP [Nation I., 2001; 64]. Chung and Nation suggested that there are considerable researches about high frequency and academic words, but findings showed that there has been a few about specialized vocabulary. Nonetheless, specialized vocabulary is widely used by people

working or studying in a certain field which plays a big role in ESP learning. According to Chung & Nation specialized words are considered as a special group of vocabulary as they are low frequency words limited to a certain area in which they appear with a fairly high frequency [Nation I., 2001; 64; Nation, P., Hwang K., 1995 35; Chung, T., Nation, P., 2004; 254; Pearson, Z. 1998; 347]. Specialized vocabulary exists in a scope of a certain subject, is found in a special field and is part of the subject knowledge system [Chung, T., Nation, P., 2004; 252]. Nation defined specialized vocabulary as words that are “recognizably specific to a particular topic, field or discipline” [6. P. 198]. Acquiring specialized vocabulary is crucial to learners to attain academic literacy and to become part of their chosen academic discourse communities. According to Waring and Nation [16. Vol.4. P. 97], a reader needs to know at least 95% of the surrounding vocabulary to successfully guess the meaning of a new word through context. They claim that ‘if readers are not capable of doing so, then they will need to interrupt their reading to find the meaning of the word in a dictionary. As Chung and Nation [Chung, T., Nation, P., 2004; 252] observed, there is a need for further research focused on specialized vocabulary words because the number of the words and their characteristics are not well understood.

## 1.2. Collocations

When one states about a collocation he certainly will define it by the most cited note done by one of the most prominent figures in British Linguistics Firth, “You shall know a word by the company it keeps!” [Firth J., 1957; 4].

The maxim is used in order to draw our attention to the fact that in natural language, words are not randomly combined, constrained only by syntax, but they have preferences. However, Firth’s definition on a collocation is not clear enough. This in turn forced a number of scholars to clarify it, consequently there appeared a multitude of different definitions, based on the views of researchers and the particular approaches on which collocations were to be applied, i.e. there does not exist general agreement on the definition of a collocation. As a consequence, there were coined a large number of terms alternative to collocations that are used almost interchangeably, such as multi-word expressions (MWE), multi-word units (MWU), bigrams (or n-grams<sup>12</sup> in general), idioms and co-occurrences [Evert S., 2004; 16]. Nonetheless, for the sake of argument, we agreed with the definition given by Choueka [19. P. 610], which is quite adequate “A collocation is a sequence of two or more consecutive words, that has characteristics of a syntactic and semantic unit, and whose exact and unambiguous meaning cannot be derived directly from the meaning or connotation of its components”.

Study of collocations has been great interest issue in different fields in our life. Moreover, collocations can be seen to be a base of communicative competence because of a subset of formulaic sequences [Nation I., 2001; 65]. Collocations assist learners with the use of language [Shin, D., Nation, P., 2008; 339]. The importance of teaching collocations is also proved in the Common European Framework of Reference for Languages [Common European framework. Council of Europe, 2001]. Shin & Nation, 2012 state frequency as a criterion for determining the most relevant collocations in spoken English for second language learners. Thus, in process of teaching ESP the role of introducing collocations is significant. As a result, defining the appropriate collocations and their mastery is very crucial to an ICT specialist. Collocations are one of the most productive ways of enriching vocabulary/terminology in modern ICT. Based on a small corpus of collected research articles this work determines the most frequent collocations in ICT English.

### 1.3. Justification for the development of the ICT corpus

There is reasonable consensus that second language learners who need to use the language for challenging academic purposes, such as reading academic texts or reading technical texts and who normally have only a limited amount of time to do this, need systematic vocabulary learning instruction in addition to learning through extensive reading or incidental learning [Nation I., 2001; 65; Coady J., 1997; 225; Schmitt N., 2000; 400]. A corpus will not just provide insights into the contents but also that the results of the analyses will be claimed to be typical of the language from which the corpus was selected [Mukundan J., Menon S., 2008; 90]. The development of ICT corpus in this research study is vital as it will reveal the specialized collocations that the students should acquire. Specific educational corpora for fields is needed than ever now as the availability of language corpora to learners and teachers facilitates better learning [Mudraya O., 2006; 235]. The corpus created will not only help students but also facilitates the lecturers and instructors in the process of teaching and material design.

Size or length of corpus is an important factor of consideration. The overall size of the ESP corpus in ICT is determined as 803,294 words. But before determining the length of the corpus, certain decisions are taken such as – availability of resources, time for data collection and computerizing them. So far as time factor is concerned, the present corpus was completed within approximately 6 month. The matter of fact is that the length of a corpus is determined not by focusing on the overall length of the corpus, but focusing more on the internal structure of the corpus: the number of genres is to be included in the corpus, the length and number of individual text samples.

Genres included in the corpus were selected keeping in mind the purpose and utility of it. Our aim was to create ESP corpus and with the intention of implementing authentic material to the classroom the genre of the texts in the corpus is research articles.

Defining number of text and range is also important factor in creating a corpus. After selecting the genre, we determined how many numbers of article texts and the range of journals to be included in the corpus. There were a huge number of texts available in the language of ICT, but we were very selective in choosing them. Regarding to the article selection we considered the year of publications so we included articles that are most up-to-date (last 3 years).

Data collection is another crucial task of building a corpus. There are various ways to collect written texts for ESP corpus such as buying printed texts, use of library (with necessary permission), photocopying and scanning the texts etc. In this context, the issue of copyright is well maintained. So, we chose only e-version of open access articles that are peer reviewed and downloadable in PDF format from Sciencedirect.com

To study about various naturally occurring phenomenon on natural language text, a well structured text corpus is very much essential. The quality and structure of a corpus can directly influence on performance of various Natural Language Processing (NLP) applications. Language technology development works in different languages have been done at various levels. Requirements of modern research and development works lies on a structured and well covered ESP Corpus. We share our experience with constructing one such corpus including about 803,294 words of research articles in the field of ICT. It will provide a significant effort by serving as an important research tool for teaching ESP and NLP researchers.

The creation of ICT corpus in ESP is essential as it will reveal the specialized vocabulary that must be instilled in the students. The corpus created will not only help students but also facilitates the lecturers and instructors in the process of teaching and learning. The teaching and learning of English for ICT can be enhanced with the existence of the ICT corpus. ICT students may be exposed to the vocabulary important in their field in a more comprehensive way. With the creation of this corpus, English for ICT material developers would have an idea and guidelines on the vocabulary needed to be taken into consideration when developing a material or course book.

#### 1.4. Research question

The current study intends to demonstrate how to create ESP corpus for ICT in order to investigate the language used in this sphere from the aspects of the



specialized vocabulary, particularly collocations. Thus, the research questions addressed are as follows:

What are the most common (high frequent) collocations in the field of ICT?

## **MATERIALS AND METHODS**

### How data was acquired

We gathered research articles to make specialized corpora that are the most suitable resource as “... research articles are considered a type of authentic material to implement to ESP classroom” [Kwary D., 2018; 95]. 140 articles from 23 scholarly journals were used to compile the ICT corpus which includes 803,294 running words. The articles classified into seven subject areas of ICT: telecommunication, information security, computer engineering, software engineering, TV technology, radio technology and mobile communication. The used articles were published by Elsevier in period between 2015–2018. A concordancer to produce the key word in context (KWIC) and a word frequency table were developed by free downloaded Software AntConc (3.5.7 Windows 2018) in order to easify the access and use of the high-frequency words. When we opened the software, we were able to search for a word or several words to see the KWIC which is the position of the word together with its collocates (up to five word tokens to the left and five tokens to the right) that was useful and fast to make a word list of the words found in the sphere. An experimental peculiarity of our research is defined with that, that corpus consists of well-selected and recent scholarly research journal articles. The concordancer enables the search of particular words/collocations to determine their frequency in use. As a data source location were chosen the website sciencedirect.com because findings showed that it has most open access peer reviewed journals in the chosen spheres of ICT. In Table 1 the details of the ICT corpus are illustrated.

We used the AntConc software tools to do this research. The integrated suite of programs allows easy and fast detecting words behavior in texts. With the help of these tools we were able to see the usage of words in the articles. This software comprises programs such as: Concordance, Concordance Plot, File view, Clusters/N-grams, Collocates, Word List, and Keyword list. In this study, to select the collocations for analysis, the Word List function was run to identify frequently used collocations in the ESP corpus in the field of ICT.

To realize current research we created our own specialized corpora as the most suitable resource. We came to this decision because suitable specialized corpus for ICT spheres is unavailable. Considering size, following Bowker and Pearson [27], we endeavored to expand the corpus as large as possible so long as the criterion of

manageability remained untouched. Each article comprised roughly 5500 words. ICT disciplines selected according to the ICT areas taught in Uzbekistan technical universities.

**Table 1.** *The details of the ICT corpus*

<b>Sub-disciplines</b>	<b>Num. of Running Words</b>	<b>Num. of Research Articles Selected</b>	<b>Journals in ICT field and the total number of articles taken from each:</b>
<i>Computer Engineering</i>	117628	20	Computer Networks (2), ICT Express (5), International Journal of Electronics and Communications (5), Journal of Electrical Systems and Information Technology (8)
<i>Information security</i>	108288	20	Computer and security (1), Computer Standards & Interfaces (1), Computers in Industry (1), Energy Procedia (5), Korean Nuclear Society (1), Procedia Computer Science (8), Procedia Manufacturing (5), Transportation Research Part A: Policy and Practice (1)
<i>Mobile Communi-cation</i>	123192	20	Future Generation Computer Systems (1), ICT Express (5), International Journal of Electronics and Communications (6), Journal of Electrical Systems and Information Technology (7), Journal of King Saud University – Computer and Information Sciences (1)
<i>Radio Technology</i>	125836	20	Development Engineer (1), Digital Communications and Networks (1), Digital Investigation (1), Electronic Notes in Theoretical Computer Science (1), ICT Express (3), International Journal of Human-Computer Studies (2), Journal of King Saud University – Computer and Information Sciences (1), Procedia Computer Science (9), Transportation Research Procedia (1)
<i>Software Engineering</i>	101322	20	Egyptian Informatics Journal (2), Future Generation Computer Systems (1), ICT Express (7), Journal of Electrical Systems and Information Technology (9)
<i>Telecommu-nication</i>	111732	20	Computer Networks (1), ICT Express (4), International Journal of Electronics and Communications (10), Journal of Electrical Systems and Information Technology (5)
<i>TV technology</i>	115296	20	Ad Hoc Networks (1), Digital Communications and Networks (1), Digital Investigation (1), Discrete Applied Mathematics (6), Egyptian Informatics Journal (1), ICT Express (1), Journal of King Saud University – Computer and Information Sciences (1), Procedia Computer Science (4), Theoretical Computer Science (1)
<b>Total</b>	803294	140	23 (journals)

Since languages change over time, the language items used in the past can be different from those used nowadays. In fact, research articles in any field usually deal with new concepts. Practice shows that to interpret the data on the concepts authors use new/unknown/up-to-date terms/specialized vocabulary to define them in their work. Therefore, as a material of our corpus we chose research articles.

The articles for each section downloaded in PDF format were converted into .txt files which were uploaded into AntConc (A freeware corpus analysis toolkit for concordancing and text analysis). Then we conducted our analysis using concordance tools. The first step was to extract the most frequent content words in the corpus with (minimum frequency of three tokens). Then we determined the collocations of the frequently used key words by using collocates tool of the software. The identified collocations from each section's database were tabulated and top 10 frequency lists for each subfield are presented in Table 2.

## RESULT AND DISCUSSION

The total of 803,294 tokens (running words) were found from all the journal articles.

In this study, the low frequency collocations are defined as the collocations, which occur once up to ten times in each corpus.

**Table 2.** *The ICT Collocation Frequency List (top 10 for each subfield)*

	<b>Radio technology</b>	<b>Mobile communication</b>	<b>Computer engineering</b>	<b>Telecommunications</b>	<b>Information security</b>	<b>Software engineering</b>	<b>TV technology</b>
1	Cognitive radio (298)	Mobile communication (132)	Computer engineering (115)	Network analysis (95)	Cyber attacks (136)	Software engineering (353)	Message broadcast (115)
2	Radio networks (145)	Mobile agent (120)	Engineering education (110)	Social network (69)	Cyber (and) physical (225)	Computer science (243)	Occurs at/by time (82)
3	Cognitive network (144)	Mobile agents (66)	Computer science (100)	Complex networks (79)	Cyber risk (223)	Software systems (186)	Multi broadcast (78)
4	Radio communication (115)	Security properties (60)	Software engineering (84)	Network structure (60)	Cyber attack (187)	Software development (160)	Broadcast storm (77)
5	Phase algorithm (80)	Communication system (56)	Blended learning (61)	Biological network (58)	Cyber security (178)	Intellectual capital (134)	Dominating set (74)
6	Radio production (78)	Mobile system (50)	Modular development (43)	Data analysis (57)	Manufacturing systems (133)	Software system (97)	Source node (65)
7	Communication equipment (54)	Core network (45)	User interface (41)	Network data (52)	Attack path (79)	Social capital (88)	Broadcast problem (64)
8	Distributed algorithm (49)	Mobile satellite (43)	Systems engineering (40)	World network (46)	Cyber securities (79)	Cloud computing (820)	Broadcast algorithm (59)
9	Wireless communication (48)	Access control (32)	User experience (34)	Network topology (45)	Risk assessment (77)	Software design (66)	Broadcast time (58)
10	Radio system	Mobile	User	Time data	Security	Human capital	Time

(45)	technologies (29)	interfaces (33)	(42)	system (69)	(64)	unit/unit of time (53)
------	----------------------	-----------------	------	-------------	------	---------------------------

The 10 most common collocations used in journal articles in the field of ICT are identified after removing all the functional words in the English language such as articles, pronouns and other functional grammatical items in the language.

The ICT Collocation Frequency List:

1. Software Engineering (353)
2. Cognitive radio (298)
3. Computer science (243)
4. Cyber (and) physical attack (225)
5. Cyber risk (223)
6. Cyberattack (187)
7. Software systems (186)
8. Cyber security (178)
9. Software development (160)
10. Radio networks (145)

In this paper, we presented a description of processes involved in defining high frequently used collocations in ICT through ESP corpus of ICT research articles. Corpus is being regarded as a multi-dimensional in nature.

## **CONCLUSION**

Corpus opens up new avenues in the field of language technology, communication, exchange of information, translation, language education and linguistic activities etc. Teaching and learning English for ICT purposes can be enhanced with the existence of the ICT corpus. With the creation of this corpus, material developers of English for ICT purposes would have an idea and guidelines on the vocabulary needed to be taken into consideration when developing a material or an English for ICT purposes course book (separate to each sub-field). On the other hand, the production of the ICT dictionaries can be compiled not only based on the general ICT terminologies but also focusing on the frequently used collocations in each ICT sub-fields: telecommunication, information security, computer engineering, software engineering, TV technology, radio technology and mobile communication. We hope that in the future a large scale of corpora for ICT will be created. Besides, steps are to be taken in annotating the raw corpus which would result in building morphological analyzer, spell checking tool, concordancer, machine translation, speech recognition etc. in the language of ICT industry. In our further research we are going to suggest how to implement chosen collocations in a language classroom.

Text mining and the use of corpora help in systematizing linguistic knowledge, automating the text processing process in order to define specialized lexical units in various spheres. The largest percentage of existing corpora belongs to the educational segment, thus, the linguistic corpus is a means for solving not only scientific, but also educational, methodological and lexicological problems. The benefits of its application in various fields are beyond doubt.

## REFERENCES

1. Abedi, Z. Mobaraki, M. (2014) The Effect of Grammatical Collocation Instruction on Understanding ESP Texts for Undergraduate Computer Engineering Students. *Language Teaching and Research*, 5(3), 631-641.
2. Özdemir, N.Ö. (2014) Using Corpus Data to Teach Collocations in Medical English. *Journal of Second Language Teaching and Research*. 3(1), 37-52.
3. Mohamad, A.F., Ng, Yo Jin. (2013) Corpus-based studies on Nursing Textbooks. *Advances in Language and Literary Studies*, 4(2), 21-27.
4. Chung, T., & Nation, I. S. P. (2003). Technical Vocabulary in Specialized Texts. *Reading in a Foreign Language*, 15, 103-116.
5. Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213–238.
6. Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge, England: Cambridge University Press. 60-113.
7. Nation, P., Waring, R. (1997) In Schmitt, N. and M. McCarthy (Eds.): *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press. 6-19.
8. McEnery, T. & A. Wilson (2001). *Corpus Linguistics* (2nd ed.). Edinburgh: Edinburgh University Press. 235 p.
9. Kadirbekova, D. (2017). *English-Uzbek Terminology of ICT and its Linguistic Peculiarities*. PhD diss. in phil. scien. pp. 80-82. Tashkent
10. L'Homme, M. (2009). A Methodology for Describing Collocations in a Specialized Dictionary. In *Collocations in Specialized Dictionaries*, p. 2.
11. Hutchinson, T., & Waters, A. (1987). *English for Specific Purposes: A Learner-Centered Approach*. Cambridge: Cambridge University Press.  
<http://dx.doi.org/10.1017/CBO9780511733031>
12. Dudley-Evans, T., St. John, M. J. (1998). *Developments in English for Specific Purposes*. Cambridge: Cambridge University Press.
13. Nation, P., Hwang K. (1995). Where would general service vocabulary stop and special purposes vocabulary begin? *System*, 23, 35-41.
14. Chung, T.M., Nation, P. (2004), Identifying technical vocabulary. *System*, 32, 143-300.

15. Pearson, Z. B. (1998). Assessing Lexical Development in Bilingual Babies and Toddlers. *The International Journal of Bilingualism*, 2(3), 347-372.
16. Waring, R., Nation, P. (2004). Second Language Reading and Incidental Vocabulary Learning. *Angles on the English Speaking World*, 4, 97-110.
17. Firth, J.R. (1957). 'A synopsis of linguistic theory, 1930-1955'. *Studies in Linguistic Analysis. Special volume of the Philological Society*. Oxford: Blackwell. 1-32.
18. Evert, S. (2004). The Statistics of Word Cooccurrences: Word Pairs and Collocations. *PhD diss., IMS*, pp. 18-20. University of Stuttgart.
19. Choueka, Y. (1988). Looking for needles in a haystack. *Proceedings of RIAO*, 609-623.
20. Shin, D., Nation, P. (2008). Beyond single words: The most frequent collocations in spoken English. *ELT Journal*, 62 (4), 339-348.
21. Common European framework. Council of Europe, 2001. [www.coe.int/t/dg4/linguistic](http://www.coe.int/t/dg4/linguistic)
22. Coady, J. (1997). L2 vocabulary acquisition through extensive reading. In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition*. 225-237.
23. Schmitt, N. (2000) Key concepts in ELT: lexical Chunks. *ELT Journal*, 54 (4), 400-401.
24. Mukundan, J., Menon, S. (2008) Nouns and their extended units of meaning: A corpus analysis of nouns used in the Science and English Language textbooks. *Jurnal Sastra Inggris*, 8 (2), 90-111.
25. Mudraya, O. (2006) Engineering English: a lexical frequency instructional model. *English for Specific Purposes*, 235-256.
26. Kwary, D. (2018) A Corpus and a Concordancer of Academic Journal Articles. *Data in brief*, 16, 94-100.
27. Bowker, L., Pearson, J. (2002) *Working with Specialized Language: A Practical Guide to Using Corpora*. London/New York: Routledge.
28. Kallas E. O., Koppel K. R., Kallas R. G. *Automatic corpus-based compilation of the collocations dictionary*. Proceedings of the international conference «Corpus linguistics–2017», p. 195.
29. Maiorova A.D. (2017). Corpus linguistics: historical and linguistic diagnostic aspects. *International research journal*, 5(59), 42-46.
30. Kallas E. O., Koppel K. R., Kallas R. G. (2017). *Automatic corpus-based compilation of the collocations dictionary*. Proceedings of the International Conference on Corpus Linguistics, p. 195.

31. Simpson, R. & Mendis D. (2003). A corpus-based study of idioms in academic speech. *TESOL Quarterly*, 27 (3), 419-441.
32. Varley, S. (2008). I'll just look that up in the concordance: Integrating corpus consultation into the language learning environment. *Computer Assisted Language Learning*, 22 (2), 133-152.