

ЛИНГВИСТИКА И МЕЖКУЛЬТУРНАЯ КОММУНИКАЦИЯ LINGUISTICS AND INTERCULTURAL COMMUNICATION

Научная статья
УДК 81'33

ПРАГМА-КОММУНИКАТИВНЫЙ АНАЛИЗ КАК ЛИНГВИСТИЧЕСКИЙ ИНСТРУМЕНТ ЖАНРОВОГО АННОТИРОВАНИЯ СПЕЦИАЛЬНОГО КОРПУСА ТЕКСТОВ

Елена Юрьевна Балашова

Саратовская государственная юридическая академия, Саратов, Россия

AuthorID: 690142, ORCID: 0000-0002-9993-8116, balashovaelena@yandex.ru

Аннотация. В статье представлена методология прагма-коммуникативного анализа текстов параллельного корпуса архивных документов ООН (документы с 1990 по 2014 гг., переведённые на шесть официальных языков ООН), направленная на создание принципов жанрового аннотирования специальных корпусов. В ряду хорошо разработанных видов корпусной разметки (просодическая, частеречная, семантическая, синтаксическая и проч.) жанровое аннотирование представляет особую сложность, как с лингвистических позиций, так и с позиций создания автоматизированных компьютерных программ. В данном случае прагма-коммуникативный анализ может выступать эффективным лингвистическим инструментом, позволяющим описать жанровый состав исследуемого корпуса текстов, а также провести его функциональный анализ. Определяются ключевые терминологические области параллельного корпуса архивных документов ООН: «Законодательство/Legislation», «Юридические документы/Legal Documents», «Аналитические материалы/Analytical materials», «Деловая корреспонденция/Business Correspondence». Проводится анализ таких базовых прагма-коммуникативных функций исследуемого корпуса текстов, как **аргументативная** (представлена всеми типами рекомендательных документов: policy recommendations, working paper (recommendations with guidance), recommendation, note, public speech, argumentative part of a report), **информативная** (репрезентирована такими жанрами, как informative report, outline of schedule, programme of meetings, provisional overview, agenda, back-ground paper, executive summary), **инструктивная** (объективирована следующими жанровыми образцами: administrative instructions, technical instructions, working paper, guidelines, decision, proposal, technical regulation), **аналитическая** (analytical part of White Paper, analytical report \ review, background, injury, legal analysis of legislation, legal research paper, legal study) и **регулятивная** (article, regulation, statute, convention, charter, resolution, Bill, court statement, decree, norms, draft, special provision). Описывается определённый инвентарь интратекстуальных свойств корпусных текстов, включающий их лексические и морфолого-синтаксические характеристики. Делаются выводы о возможности применения предлагаемой методологии к жанровому аннотированию различных видов корпусов.

Ключевые слова: специальный корпус текстов, корпусная разметка, жанровое аннотирование, прагма-коммуникативные функции, интратекстуальные свойства, терминологическая область, юридические жанры.

Original article

PRAGMA-COMMUNICATIVE ANALYSIS AS A LINGUISTIC TOOL OF GENRE ANNOTATION OF A SPECIALISED TEXT CORPUS

Elena Y. Balashova

Saratov State Law Academy, Saratov, Russia

AuthorID: 690142, ORCID: 0000-0002-9993-8116, balashovaelena@yandex.ru

Abstract. The article presents a methodology of pragma-communicative analysis of archival texts from the UN Parallel Corpus (documents from 1990 to 2014, translated into the six official UN languages), aimed at creating the principles of genre annotation of specialised corpora. Among the well-developed types of corpus annotation (prosodic, part-of-speech tagging, semantic, syntactic, etc.), genre annotation is particularly difficult, both from the perspective of linguistics and the perspective of creating automated computer programmes. In this case, pragma-communicative analysis can act as an effective linguistic tool that allows describing the genre composition of the studied corpus of texts, as well as conducting its functional analysis. The key terminological areas of the UN Parallel Corpus are defined as: «Законодательство/Legislation», «Юридические документы/Legal Documents», «Аналитические материалы/Analytical materials», «Деловая корреспонденция/Business Correspondence». The article contains analysis of such basic pragma-communicative functions of the studied corpus of texts as **argumentative** (represented by all types of recommendation documents: policy recommendations, working paper (recommendations with guidance), recommendation, note, public speech, argumentative part of a report), **informative** (represented by such genres as informative report, outline of schedule, programme of meetings, provisional overview, agenda, back-ground paper, executive summary), **instructional** (objectified by the following genre samples: administrative instructions, technical instructions, working paper, guidelines, decision, proposal, technical regulation), **analytical** (analytical part of White Paper, analytical report\ review, back-ground, injury, legal analysis of legislation, legal research paper, legal study) and **regulatory** (article, regulation, statute, convention, charter, charter, resolution, Bill, court statement, decree, norms, draft, special provision). A certain inventory of intratextual properties of corpus texts is described, including their lexical and morphological-syntactic characteristics. The article is concluded with possibilities of applying the proposed methodology to genre annotation of various types of corpora.

Keywords: special corpus of texts, corpus markup, genre annotation, pragma-communicative functions, intratextual properties, terminological domain, legal genres.

Введение и постановка проблемы. В современной корпусной лингвистике существует ряд разработанных методик как экстралингвистической (специфика форматирования текста, сведения о тексте и его авторе, язык текста, выходные данные издания), так и лингвистической разметки (частеречная, семантическая, просодическая, синтаксическая, анафорическая) [см. Захаров, Хохлова, 2015; Захаров, 2014; Kirk, 2015]. Однако жанровая разметка остаётся одной из самых сложных задач современной компьютерной лингвистики, а между тем подобный тип разметки крайне важен для специальных корпусов или корпусов подязыка, как правило содержащих относительно ограниченный набор жанров и специфичную терминологию [см. Дементьев, Степанова, 2016; Lehberger, 1986].

Прагма-коммуникативный функциональный анализ текстов специального корпуса, на наш взгляд, выступает эффективным лингвистическим инструментом разработки принципов жанрового аннотирования. Подобный анализ позволяет выделить набор базовых прагма-коммуникативных функций, реализуемых теми или иными жанрами корпусных текстов, описать интратекстуальные характеристики основных жанровых образцов и выявить устойчивые морфолого-синтаксические паттерны специального корпуса текстов.

S. Sharoff в работе «Functional Text Dimensions for the Annotation of Web Corpora» [Sharoff, 2018: 25] выделяет следующие базовые функциональные измерения текстов, составляющих корпуса общезыкового неспециального характера (веб-корпус UkWas, а также традиционные корпуса British National Corpus и the Brown Corpus):

- | | |
|---|--------------------------------------|
| 1. аргументативное \ argum (A1); | 9. персональное \ personal (A11); |
| 2. эмотивное \ emotive (A3); | 10. рекламное \ compuff (A12); |
| 3. реальное \ фиктивное \ fictive (A4); | 11. идеологическое \ ideopuff (A13); |
| 4. развлекательное \ flippant (A5); | 12. аналитическое \ scitech (A14); |
| 5. неформальное \ informal (A6); | 13. специальное \ specialist (A15); |
| 6. инструктивное \ instruct (A7); | 14. информативное \ info (A16); |
| 7. новостное \ hardnews (A8); | 15. оценочное \ eval (A17); |
| 8. регулятивное \ legal (A9); | 16. диалогическое \ dialogue (A18); |

17. поэтическое \ poetic (A19);

18. апеллятивное \ appell (A20).

Выделенные функциональные измерения реализуют определённые прагма-коммуникативные функции корпусных текстов в современной коммуникации. Так, аргументативная функция представлена измерениями A1, A13, A17, информативная – измерениями A8, A16, инструктивная функция репрезентирована измерениями A7, A9, A20, аналитическая – измерением A14 и, наконец, регулятивная прагма-коммуникативная функция реализована измерением A9.

Методология исследования. Мы полагаем, что набор функциональных измерений и реализуемых ими прагма-коммуникативных функций не столь разнообразен в специальном корпусе, нежели в корпусах общезыкового характера. Это объясняется меньшим разнообразием жанрового состава специального корпуса. Так, параллельный корпус архивных документов ООН (документы с 1990 по 2014 гг., переведённые на 6 официальных языков ООН: арабский, китайский, английский, французский, русский, испанский – перевод ручной) [Ziemski, 2016: 3530] содержит такие крупные терминологические области, как «**Законодательство** \ *Legislation*», «**Юридическая документация** \ *Legal Documentation*», «**Правовой сопоставительный анализ** \ *Legal research comparative analysis*» и «**Деловая корреспонденция** \ *Business Correspondence*». К терминологической области «**Legislation**» принадлежат такие жанры, как *постановления, указы, распоряжения, приказы, резолюции, декларации, хартии, статуты, акты, конвенции*. Терминологическая область «**Legal Documentation**» представлена *контрактами, международными соглашениями, информационными бюллетенями, протоколами, меморандумами, отчётами*, а также таким довольно редким законодательным жанром, как *Белая книга* (white paper). Область «**Legal research comparative analysis**» включает следующие жанры корпусных текстов: *аналитические обзоры, аналитические комментарии к законодательным документам, исследовательские отчёты*. Наконец, терминологическая область «**Business Correspondence**» содержит *дипломатические и правительственные письма и циркуляры* [см. Berūkštienė, 2016; Swales, 1990].

Представляется целесообразным провести прагма-коммуникативный анализ жанровых образцов исследуемого корпуса текстов.

1. Аргументативная функция (все типы рекомендательных документов: policy recommendations, working paper (recommendations with guidance), recommendation, note, public speech, argumentative part of a report):

- <...> *For the purposes of this analysis, two cases may be of particular relevance: the Klass case and the Leander case. In the Klass case concerning secret surveillance, the ECHR (the European Court of Human Rights) held that the remedy should be as effective as could be having regard to the restricted scope for recourse inherent in any system of secret surveillance. In the Leander case concerning security checks, the court further held that even if no single remedy under the national system might be effective on its own, the aggregate of remedies as a whole might still qualify as effective. More specific guidelines as to what constitutes an effective remedy might be found in a recommendation of the Committee of Ministers of the Council of Europe, which specifies the right to an effective remedy for rejected asylum seekers against decisions on expulsion. In Recommendation R (98) 13, the committee held that a remedy before a national authority is considered effective when: That authority is judicial; That authority has competence both to decide on the existence of the conditions [...] and to grant appropriate relief; The remedy is accessible for the rejected asylum seeker. Although of course not directly applicable, this recommendation may still provide useful guidance as to what is required under Article 13 of the ECHR <...>*

2. Информативная функция (informative report, outline of schedule, programme of meetings, provisional overview, agenda, back-ground paper, executive summary, follow-up review, periodic briefings, work programme):

- <...> *Notification of an individual or entity of their inclusion on sanctions lists can only take place to the extent practicable and is not a prerequisite for listing. In cases of renowned terrorists or individuals evading international law enforcement, the location of targets is not known, making notification impossible. <...> Procedural issues like the norms concerning identifying information for listed individuals and entities, sanctions committees have developed general norms about handling exemptions. The grounds for exemptions and information required for exemptions are increasingly standardized across the different committees, as there appears to have been much learning and borrowing of best practices from other committees in recent*

years. This is again indicative of improvements in committee practices since targeted sanctions were first introduced in the early 1990s. Exemptions for basic needs also help ensure that proportionality standards are met and that their application does not inappropriately interfere with fundamental rights <...>;

- The Committee is of the opinion that the revised estimates for the period from 1 May to 31 July 1993 and the estimates for the period from 1 to 31 August 1993 are not complete and that additional information needs to be provided to the Committee for consideration. Upon inquiry, the Committee was informed that the estimates for both military and civilian personnel are based on their full deployment as at 1 May 1993. The Committee is of the view that there is a need to indicate the planned deployment and actual date of arrival of all additional military and civilian personnel, delays in deployment (person-days), vacancy rates and related savings;

3. Инструктивная функция (administrative instructions, technical instructions, working paper, guidelines, decision, proposal, technical regulation, directive, ordinance):

- In order to ensure receipt of all applications submitted by staff members from offices away from Headquarters, such staff members are requested to return their applications to the Strategic Planning and Staffing Division of the Office of Human Resources Management through the Chief Administrative Officer or Director of the United Nations Information Centre of their respective duty stations by 20 March 2009, the deadline for receipt of applications in the Examinations and Tests Section.

4. Аналитическая функция (analytical part of White Paper, analytical report \ review, back-ground, injury, legal analysis of legislation, legal research paper, legal study):

- In sum, this section demonstrates that various human rights may be compromised where targeted sanctions are concerned. Whether certain sanctions actually infringe upon specific human rights depends largely on the kind of sanctions involved. The most problematic issues concern procedures that are followed in applying targeted sanctions. As regards those procedural aspects, the main issues in applying financial sanctions and travel bans concern listing and delisting procedures, the procedural system of granting exemptions, and whether there is periodic review of the effectiveness and necessity of the sanctions imposed. The right to a fair trial would require judicial review of the sanctions, but we can ask whether that right is applicable to the administrative procedures by which UN sanctions are imposed. In contrast, it is likely that the right to an effective remedy is applicable. This right leaves more room for different kinds of remedies, as long as they are as effective as possible in the context of the situation in which they apply.

5. Регулятивная функция (article, regulation, statute, convention, charter, resolution, Bill, court statement, decree, norms, draft, special provision):

- Enhancing the coordinating functions of central authorities designated under article 18, paragraph 13, of the Organized Crime Convention Article 18, paragraph 13, of the Organized Crime Convention requires States parties to designate a central authority with the responsibility and power to receive requests for mutual legal assistance and either to execute them or to transmit them to the competent authorities for execution. A similar obligation is contained in article 7, paragraph 8, of the United Nations Convention against Illicit Traffic in Narcotic Drugs and Psychotropic Substances of 1988 and in article 46, paragraph 13, of the Convention against Corruption. In addition to the requirement of article 18, paragraph 13, of the Organized Crime Convention, States parties to the Protocol against the Smuggling of Migrants by Land, Sea and Air, supplementing the Organized Crime Convention, are required to designate an authority or authorities to receive and respond to requests relating to vessels (art. 8, para. 6, of that Protocol). States parties to the Protocol against the Illicit Manufacturing of and Trafficking in Firearms, Their Parts and Components and Ammunition, supplementing the Organized Crime Convention, must also identify a national body or single point of contact to act as a liaison on matters relating to the Protocol (art. 13, para. 2, of that Protocol).

Результаты и обсуждение. Представляется целесообразным описать инtratекстуальные характеристики основных жанровых образцов, репрезентирующих перечисленные прагма-коммуникативные функции. К инtratекстуальным характеристикам относятся такие лингвистические свойства жанров, как морфолого-синтаксические и лексические особенности [см. Sharoff, 2021].

Так, жанровые образцы, которые реализуют аргументативную, информативную, инструктивную, аналитическую и регулятивную прагма-коммуникативные функции в исследуемом специальном корпусе юридических текстов обладают следующими **морфолого-синтаксическими особенностями**:

1) наличие сложных предложений с бессоюзной связью, осложнённых однородными членами: «...two cases may be of particular importance: *the Klass case and the Leander case*...»; «...the committee held that a remedy before a national authority is considered effective when: *That authority is judicial; That authority has competence both to decide on the existence of the conditions and to grant appropriate relief; The remedy is accessible for the rejected asylum seeker*...»;

2) сложные предложения с подчинительной связью, осложнённые причастными оборотами с функцией уточнения: «In the Klass case *concerning secret surveillance*, the European Court of Human Rights held that...»; «...the remedy should be as effective as could be *having regard to* the restricted scope for recourse inherent in any system of secret surveillance...»;

3) сложноподчинённые предложения с обособленными осложняющими конструкциями: «...*For the purpose of this analysis*, two cases may be of particular relevance...»; «...*Although of course not directly applicable*, this recommendation may still provide useful guidance...»; «...procedural issues *like the norms concerning identifying information for listed individuals and entities*, sanctions committees have developed general norms...»; «...*As regards those procedural aspects*, the main issues in applying financial sanctions and travel bans concern listing and delisting procedures...»; «...*Upon inquiry*, the committee was informed that...»;

4) наличие обособленных союзных слов, выполняющих функцию логической связи и организации частей сложного предложения (данная особенность характерна для жанров, репрезентирующих аналитическую прагма-коммуникативную функцию): «...*In sum*, this section demonstrates that various human rights may be compromised...»; «*In contrast*, it is likely that the right to an effective remedy is applicable...»;

5) инфинитивные обороты и конструкции для выражения цели, свойственные жанрам, представляющим инструктивную функцию: «...*in order to* ensure receipt of all applications submitted by staff members...»;

6) наличие составных подчинительных союзов с функцией структурной логической организации текста: «...the remedy should be *as* effective *as* could be...»; «...competence *both* to decide on the existence of the conditions *and* to grant appropriate relief...»; «...This right leaves more room for different kinds of remedies, *as long as* they are *as* effective *as* possible in the context of the situation in which they apply...»;

7) на текстовом уровне юридическим жанрам исследуемого корпуса свойственны такие приёмы логической организации текста, как нумерация однородных членов (или однородных придаточных предложений), а также парентеза: «The Committee is of the view that there is a need to indicate...delays in deployment (*person-days*) ...».

Лексические особенности жанров, принадлежащих корпусу архивных документов ООН:

1) суазивные глаголы со значением убеждения (require, request, encourage, recommend, propose, call upon, offer, suggest, urge, appeal, ask и т.п.): «...we can *ask* whether that right is applicable to the administrative procedures...»; «...staff members are *requested*...»; «...the right to a fair trial would *require* judicial review of the sanctions...»;

2) модальные глаголы со значением вероятности: «...the aggregate of remedies as a whole *might* still qualify as effective...»; «...two cases *may* be of particular importance»; «...an effective remedy *might* be found in a recommendation»;

3) официальные клише и обороты: «...*designated under article*...»; «...*be of particular relevance*...»; «...*the court held*...»; «...*upon inquiry*»;

4) специальные юридические термины и терминологические сочетания: «*execution*»; «*the right to a fair trial*»; «*judicial review*»; «*remedy*»;

5) модальные глаголы со значением необходимости: «...the remedy *should* be as effective as could be...»; «states parties...*must* also identify a national body...»;

6) модальные глаголы со значением возможности: «notification of an individual or entity of their inclusion on sanctions lists *can* only take place».

Выводы и перспективы дальнейшей исследовательской работы. Данные прагма-коммуникативного анализа могут быть использованы для жанрового аннотирования специальных корпусов

текстов, поскольку позволяют определить набор базовых жанров того или иного корпуса, описать специфичные функции, реализуемые выделенными жанровыми образцами, а также составить своеобразный инвентарь морфолого-синтаксических и лексических особенностей ключевых жанров специального корпуса, что необходимо для разработки принципов автоматизированной жанровой разметки электронных корпусов текстов.

Как показал анализ, специальный корпус (UN Parallel Corpus) репрезентирован немногочисленным составом прагма-коммуникативных функций (5 базовых функций: аргументативная, информативная, инструктивная, аналитическая, регулятивная) в отличие от корпуса общезыкового характера (например, ukWac), представленного 18 прагма-коммуникативными функциями. Мы полагаем, что специальным корпусам текстов (корпусам юридических, технических и проч. текстов) присущ достаточно ограниченный набор прагма-коммуникативных функций в силу их характерной терминологии и относительно небогатого жанрового состава.

Методика прагма-коммуникативного анализа жанровых образцов, принадлежащих тому или иному корпусу, может быть использована для разработки различных видов аннотирования как специальных, так и неспециальных корпусов текстов. Основной задачей подобного рода исследований является определение жанрового состава изучаемого корпуса, а также описание его базовых функциональных и интратекстуальных характеристик.

Список источников

Дементьев, В. В. Корпусные методы в исследовании речевых жанров / В. В. Дементьев, Н. Б., Степанова // Вестник РУДН Серия: Лингвистика. 2016. Т. 20. № 3. С. 57-76.

Захаров, В. П. Автоматическое выявление терминологических словосочетаний / В. П. Захаров, М. В. Хохлова // Структурная и прикладная лингвистика. 2014. № 10. С. 182-200.

Захаров, В. П. Корпусный подход к построению тезауруса и онтологии / В. П. Захаров // Структурная и прикладная лингвистика. 2015. № 11. С. 123-143.

Berūkštienė, D. (2016) Legal discourse reconsidered: genres on legal texts. *Comparative Linguistics* 28: 89-117.

Ziemski, M., Junczys-Dowmunt, M., Pouliquen, B. (2016) *The United Nations parallel corpus v1.0. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association (ELRA): 3530-3534.

Kirk, J. (2015) Corpus annotation: Speech acts and the description of spoken registers. *Corpus Linguistics. Abstract Book* ed. by F. Formato and A. Hardie. Lancaster: UCREL: 197-199.

Lehberger, J. (1986) *Sublanguage analysis. Analyzing language in restricted domains: sublanguage description and processing*. New Jersey: Lawrence Erlbaum Associates: 19-38.

Sharoff, S. (2018) Functional Text Dimensions for annotation of Web corpora. *Corpora* 13 (1): 65-95.

Sharoff, S. (2021) Genre annotation for the Web: text-external and text-internal perspectives. *Register Studies* 3, 1: 1-32.

Swales, J. (1990) *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press: 260.

References

Berūkštienė, D. (2016) Legal discourse reconsidered: genres on legal texts. *Comparative Linguistics*, 28: 89-117.

Dementiev, V. V., Stepanova, N. B. (2016). Korpusnyje metody v issledovanii rechevykh zhanrov [Corpus methods of speech genres]. *Vestnik RUDN. Seriya: Lingvistika* [Bulletin of the Peoples' Friendship Russian University. Linguistics], (20) 3: 57-76. (In Russ.)

Kirk, J. (2015) Corpus annotation: Speech acts and the description of spoken registers. *Corpus Linguistics. Abstract Book* ed. by F. Formato and A. Hardie. Lancaster, UCREL: 197-199.

Lehberger, J. (1986) *Sublanguage analysis. Analyzing language in restricted domains: sublanguage description and processing*. New Jersey, Lawrence Erlbaum Associates: 19-38.

Sharoff, S. (2018) Functional Text Dimensions for annotation of Web corpora. *Corpora* 13 (1): 65-95.

Sharoff, S. (2021) Genre annotation for the Web: text-external and text-internal perspectives. *Register Studies* 3, 1: 1-32.

Swales, J. (1990) *Genre analysis: English in academic and research settings*. Cambridge, Cambridge University Press: 260.

Zakharov, V. P., Khohlova, M. V. (2014). Avtomaticheskoye vyjavleniye terminologicheskikh slovosochetaniy [Automatic identification of terminological phrases]. *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics], 10: 182-200. (In Russ.)

Zakharov, V. P. (2015). Korpusnyj podkhod k postroeniyu tezaurus i ontologii [Corpus-Based Approach to Thesaurus and Ontology Construction]. *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics], 11: 123-143. (In Russ.)

Ziemski, M., Junczys-Dowmunt, M., Pouliquen, B. (2016) *The United Nations parallel corpus v1.0. Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portorož, Slovenia: European Language Resources Association (ELRA): 3530-3534.

Информация об авторе

Елена Юрьевна Балашова – доктор филологических наук, профессор кафедры иностранных языков, Саратовская государственная юридическая академия, ул. Вольская, 1, Саратов, Россия, 410028. Автор более 100 научных публикаций. Сфера научных интересов: теория перевода, терминология, корпусная лингвистика, генристика, дискурсология.

Вклад автора: разработка концепции, проведение исследования, подготовка и редактирование текста, утверждение окончательного варианта.

О конфликте интересов, связанном с данной публикацией, не сообщалось.

Поступила 15.12.2023; принята после рецензирования 27.12.2023; опубликована онлайн 01.02.2024.

Information about the author

Elena Y. Balashova – Doctor of Philological Sciences, Professor, Department of Foreign Languages, Saratov State Law Academy, 1 Volskaya St., Saratov, Russia, 410028. Author of more than 100 scientific publications. Research interests: translation theory, terminology, corpus linguistics, genristics, discourse studies.

Author's contribution: conceptualisation, research, drafting and editing of the text, approval of the final version.

No conflicts of interest related to this publication were reported.

Received 15.12.2023; adopted after review 27.12.2023; published online 01.02.2024.