

A corpus-based study of keywords and key terms in English used in the automotive industry and their usefulness for vocational secondary school EFL students

Justyna Kendik-Gut¹ & Łukasz Grabowski²

University of Silesia (Poland)¹, University of Opole (Poland)²

justyna.kendik-gut@us.edu.pl, lukasz.grabowski@uni.opole.pl

Abstract

The lack of up-to-date authentic teaching materials is one of the problems in teaching English for Specific Purposes (ESP) at Polish secondary vocational schools. This corpus-based study, focusing on English language used in various automotive contexts, is intended to fill this gap and it has two goals. First, it aims to identify and describe discoursal functions of keywords and key terms found in a purpose-designed 1-million-word corpus comprising car reviews, car brochures, scripts from a motoring TV programme and excerpts from car owner's manuals concerning safety. Second, by using questionnaires administered to teachers and calculating inter-rater agreement, it aims to identify those keywords and key terms that are the most pedagogically useful. The study findings reveal a high degree of linguistic variation within the four text types found in the corpus of English automotive texts and provide a list of pedagogically-useful vocabulary items that may be used in teaching ESP to future car mechanics in vocational schools or to employees in the automotive industry at large.

Keywords: corpus linguistics, automotive discourse, keywords, key terms, vocational education, English for Specific Purposes.

Resumen

Un estudio de corpus de palabras clave y términos clave en inglés utilizados en la industria automotriz y su utilidad para estudiantes de secundaria profesional de inglés como lengua extranjera

La falta de materiales didácticos auténticos y actualizados es uno de los problemas en la enseñanza del inglés con fines específicos (ESP) en las escuelas secundarias de formación profesional en Polonia. Este estudio de corpus, que se centra en el inglés utilizado en diversos contextos automotrices, tiene como objetivo cubrir esta carencia y persigue dos objetivos. En primer lugar, pretende identificar y describir las funciones discursivas de las palabras clave y los términos clave encontrados en un corpus de un millón de palabras creado a partir de reseñas de automóviles, folletos de automóviles, guiones de un programa de televisión sobre automovilismo y extractos de manuales de usuario de automóviles relacionados con la seguridad. En segundo lugar, mediante el uso de cuestionarios administrados a docentes y el cálculo del grado de acuerdo entre evaluadores, busca identificar aquellas palabras clave y términos clave que resultan más útiles desde el punto de vista pedagógico.

Los resultados del estudio revelan un alto grado de variación lingüística entre los cuatro tipos de textos que conforman el corpus de textos automotrices en inglés, y proporcionan una lista de términos de vocabulario de utilidad pedagógica que pueden emplearse en la enseñanza del inglés con fines específicos a futuros mecánicos de automóviles en escuelas de formación profesional o a empleados de la industria automotriz en general.

Palabras clave: lingüística de corpus, discurso automotriz, palabras clave, términos clave, formación profesional, inglés con fines específicos.

1. Introduction

Nowadays vocational education and training is getting more and more attention from state authorities, policy makers and educational think tanks all over the world. For example, in Poland, the report published by the Foundation for the Development of the Education System (Grzesiak, 2021) states that after completing education, graduates of Polish secondary vocational schools, including car mechanics, often decide to work outside Poland, where knowledge of foreign languages, especially English, is a *sine qua non* condition for an employee to succeed. Lonsmann and Kraft (2017) argue that in terms of the workplace environment, English is presently considered “the legitimate, or even required, lingua franca for transnational communication in blue-collar settings” (section: International Blue-Collar Workplaces: Inclusion and Exclusion). The knowledge of English may considerably affect the workers’ career prospects, as the lack of it may be used as a justification for underpaid jobs. Also, competence in English may help migrant workers to become more successful in their workplace (Lonsmann & Kraft, 2017).

Thus, we may reasonably argue that car mechanics with a suitable level of foreign language skills allowing them to communicate with colleagues, customers and to use specialist tools and equipment, would be more desirable in the job market both in Poland and abroad, also considering the large automotive industry sector in the Polish and European economy. However, the only time in Poland when future car mechanics can learn ESP in a formal setting is during English classes at secondary vocational schools, as they do not learn trades or any other subjects in English. In vocational school, their language education comprises general English lessons and ESP lessons, and English is taught as a foreign language to them. According to the timetable for the implementation of the core curriculum for vocational schools, only 30 hours in the entire process of secondary education are intended for learning a specialized, professional language. That is why in this study, we aim to develop a comprehensive description of English automotive language that may be used by ESP teachers from vocational schools to prepare teaching aids, in particular for teaching specialized vocabulary and phraseology. As Coxhead (2018) notices, research into specialized vocabulary and phraseology has been mainly focused on tertiary level education, while less focus has been put on secondary school contexts, and this provides the main motivation behind our study.

The report issued by the Foundation for the Development of the Education System (Grzesiak, 2021) also states that English language teachers in Poland experience difficulties teaching professional language, and finding one suitable classroom textbook is problematic, especially for less popular professions. ESP teachers have to develop not only the teaching materials, but also the syllabus itself. A research study among the ESP teachers in Polish vocational schools conducted by Sowa (2022) revealed that the teachers lack support in areas such as development of background specialized knowledge of a given profession, access to textbooks for different professions or scarcity of teaching materials. Considering the rapid technological advances in the field of the automotive industry and the lack of support from experts, the scope of lesson topics is not always up-to-date. Moreover, ESP teachers in vocational schools do not have a fixed framework of what vocabulary and phraseology should be taught. According to the Ministerial Decree about the core curriculum for ESP issued by the Ministry of National Education in Poland,¹ the students in secondary schools being trained to become car mechanics should reach an A2 or B1 level of language competence, according to the Common European Framework of Reference

for Languages. The core curriculum, however, does not provide information on what specific language structures should be acquired. The description of qualifications for the profession of car mechanics (Ministry of National Education in Poland, 2019),² describes learning outcomes for the subject of foreign specialized language, which include such matters as using a basic-level language enabling the implementation of professional activities, or understanding and creating simple spoken and written messages sufficient to carry out professional tasks. Moreover, students should participate actively in professional conversations and writing. But teachers must decide for themselves what might be useful for their learners. Therefore, a systematic and detailed description of English automotive language could serve as a useful starting point for developing suitable teaching materials for future car mechanics trained in secondary vocational schools.

The English automotive discourse community includes engineers, car designers, car mechanics, businesspeople, journalists, and - last but not least - car users. The members of this discourse community communicate and share information in various ways: they write and read blogs, watch and record vlogs, interact with others on Internet fora, watch advertisements and dedicated TV programmes, follow Internet channels and social media profiles devoted to cars, etc. The vocabulary, terminology, abbreviations and acronyms found there are often technical and may seem opaque, ambiguous or simply unfamiliar to people from outside the community: “torque-rotational force”, “FWD”, “CVT”, “hatch”, “oversteer”, to name but a few. Car mechanics trained in secondary vocational schools gather their knowledge of specialized English language through formal education as well as from the variety of sources mentioned above.

This study focuses on exploration of salient vocabulary and phraseology, that is, keywords and key terms, typical of the automotive industry extracted using corpus tools, and it complements earlier corpus-based or corpus-driven studies conducted on other specialized discourses, e.g., general academic (Biber et al., 2004; Chang, 2014; Simpson-Vlach & Ellis, 2010), social sciences (Boreus & Bergstrom, 2017), engineering (Fox & Tighelaar, 2015; Gilmore & Millar, 2018), legal (Goźdź-Roszkowski, 2021) or pharmaceutical (Grabowski, 2015). It also capitalizes on the results of earlier studies describing the language of the workplace (Forey, 2004; Holmes, 2015; Koester, 2004, 2006, 2010), as the study focus is on secondary vocational education preparing future car mechanics in Poland.

Anecdotal evidence suggests that secondary school students trained as car mechanics require assistance in the development of their up-to-date and relevant professional language taught in the Polish EFL context, which poses challenges (Grzesiak, 2021; Sowa, 2022) for the language instructors in trades education. What is more, the study by Coxhead et al. (2020) suggests that this particular problem not only affects Polish education but may also be familiar to ESP instructors from other countries. Parkinson et al. (2017) suggest that the learners of technical trades require an extensive vocabulary load in order to engage in technical reading and writing. Therefore, the process of learning a particular profession is a combination of developing specialized knowledge, skills, practices and literacy. As future professionals, learners have to engage in a range of interactions, including all language skills, i.e., reading, speaking, writing and listening, all of which share a common requirement, namely, being familiar with salient specialized vocabulary and phraseology including keywords and key terms.

In this study, the keywords and the key terms are extracted from a custom-designed language corpus. By keywords and key terms, we understand either single words (i.e. word forms) or two-word nominal phrases that occur in our study corpus with outstanding frequency as compared with the reference corpus (cf. Section 2 for a more detailed description). In the last two decades, specialized language corpora have become an invaluable source of data for research into vocabulary and phraseology associated with particular trades (Coxhead, 2018, 2022; Coxhead & Demecheleer, 2018; Nation, 2016), in particular focusing on the development of specialized or domain-specific word lists, e.g., engineering (Mudraya, 2006), medicine (Wang et al., 2008), nursing (Yang, 2015), business contracts (Hanks et al., 2024), or phrase lists, e.g., academic (Martinez & Schmitt, 2012), or spoken rugby discourse (Benson & Coxhead, 2022). According to Nation (2016) and Coxhead and Demecheleer (2018), wordlists have numerous benefits, as they help identify vocabulary needed by particular groups of learners, help set learning goals, and are essential in “finding out more about the nature of specialized vocabulary in the field” (Coxhead & Demecheleer, 2018, p. 84). These benefits also apply to the lists of keywords and key terms, which are technically comparable to the word lists and the phrase lists respectively.

Our analysis is based on four selected English automotive text types: car reviews retrieved from the online magazine “Top Gear”, car brochures, scripts from the motoring TV programme *Top Gear* and excerpts from car owner’s manuals concerning safety. We assume that although the above text

types generally focus on motoring, they are used in different situational and communicative contexts and targeted at different audiences. Hence, we hypothesize that the keywords and key terms identified in the course of this study are characteristic of English automotive discourse, and at the same time, that each text type prioritizes specific keywords and key terms. We argue that these lexical items will be of pedagogical value to the teachers of ESP in vocational schools in Poland and elsewhere, where ESP is taught as a foreign language, in particular for curriculum development, and designing teaching materials as well as for testing, to verify if the students learnt a pre-determined set of specialized vocabulary and phraseology, and assessment (e.g., to evaluate students' knowledge before they start the ESP course or to measure their progress of learning the target vocabulary and phraseology).

2. Methodology

2.1. Research material: the English Motor Industry Corpus (EMIC) and its sub-corpora

The study focuses on the needs of secondary-level vocational students with a specialization in car mechanics. They are 16–20-year-old students of English as a foreign language with A2/B1 language competence on the CEFR, who read - typically in their native language - specialized magazines, watch motoring TV programmes and follow online blogs or vlogs on the motor industry. We therefore analyzed four selected text types that we consider typical of automotive industry discourse: car reviews, car brochures, scripts from motoring TV programme *Top Gear* and car owners' manuals.

Car reviews (CRs) were originally published in the online magazine *Top Gear* and a random collection of 100 texts concerning passenger vehicles was extracted.³ Each review has four sections: "Overview, What is it?", "Driving, What is it like on the road?", "On the inside, Layout, finish and space", "Owning, Running costs and reliability". The communicative purpose of car reviews is primarily informational: it describes a particular car and its features, e.g., the type of car, driving comfort, design and costs of purchase and maintenance. The reviews are also meant to entertain their readers and reveal their authors' personal feelings and attitudes toward a particular car. The CRs subcorpus consists of 71,991 tokens and 60,373 types.

The collection of car brochures (CBs) consists of 50 texts concerning the 50 best-selling passenger cars in Europe in 2018, according to the website

Focus2move, which were extracted from the car manufacturers' websites.⁴ The target readers of these texts are potential buyers. Although the macro-structure of CBs differs across car manufacturers, several aspects overlap in the texts, e.g., car accessories, technological equipment, description of different parts of a car. Their communicative purpose is both informational and persuasive: apart from describing a given car in great detail, the texts aim to persuade readers into purchasing the vehicle. The CBs subcorpus consists of 361,708 tokens and 291,170 word types.

Next, 50 scripts from the motoring TV programme *Top Gear* (TGSs) were extracted from the Springfield website,⁵ which is an online database with TV show and movie scripts. The scripts under scrutiny cover 50 episodes (from season 19, episode three to season 26, episode one, including four special episodes) originally aired from January 2013 to March 2019. The addressors are the hosts of the programme, and the addressees are the audience gathered in the studio, as well as its viewers. There is an intermediate level of interaction, as the studio audience responds to the hosts with applause or laughter. The TV viewers do not interact with the addressors. The scripts represent spoken communication, which differs from the written type found in CRs and CBs, as the addressors do not have an opportunity to revise and edit their speech. The content is planned in that it follows a particular script, yet the hosts also act spontaneously during the programme. The programme can be divided into certain segments, e.g., car reviews, discussions with celebrities, races, laps around test tracks, and different types of challenges. The *Top Gear* programme's communicative purpose revolves around so-called infotainment (Boukes, 2019), that is, informing about state-of-the-art developments in the car industry in an entertaining manner (e.g., by driving a car through the desert or in sub-zero areas). The subcorpus of TGSs consists of 505,104 tokens and 386,447 word types.

The last text type under scrutiny are fragments of car owner manuals (COMs). This collection includes 20 texts dealing with car safety, and each manual represents one car brand found in the CBs subcorpus. The manuals were published between 2015 and 2020 on the car manufacturers' websites and they were extracted therefrom.⁶ The readers are typically car owners, and the texts explain various safety restrictions and describe such issues as starting the car and using safety features. The owner manuals provide procedural information and guidelines on how to perform specific actions to remain safe in the car. The subcorpus of COMs includes 212,514 tokens and 188,594 word types.

All the texts collected in the four subcorpora described above were saved in a plain text format, and together they make up a specialized corpus, which we call the English Motor Industry Corpus (EMIC) that includes 1,151,317 tokens and 926,584 word types (cf. Table 1) for make-up and size of the EMIC corpus, with statistics derived from SketchEngine (Kilgarriff et al., 2014). Although our study corpus is relatively small in size (i.e., relative to contemporary standards of large-general language corpora), it corresponds with the standards used in ESP research. For example, Koester (2010) claims that smaller corpora are more suitable than large multi-million word corpora to identify the connections between linguistic patterning and specialized contexts of language use, which is of the utmost importance in teaching ESP. The same point is made by Ghadessy et al. (2001) and Jablonkai (2022), who argue that since a specialized corpus is designed to explore a particular language variety or even a single text type or genre, it is typically smaller in size than a general language corpus. When it comes to corpus representativeness, we included four text types found in various automotive contexts. We also assumed that the topics discussed in those texts correspond to salient themes discussed in the car industry and to the interests of secondary school students trained as car mechanics. The corpus balance accounts for the compromise between the texts' size and accessibility.

Text type	No. of text samples	Size (in word tokens)
Car reviews (CRs)	100	71,991
Car brochures (CBs)	50	361,708
Top-Gear scripts (TGs)	50	505,104
Car owner's manual (COMs)	20	212,514
Total (EMIC)	220	1,151,317

Table 1. Make-up of the English Motor Industry Corpus (EMIC)

2.2. Units of analysis: keywords and key terms

As we focus on identifying salient vocabulary and phraseology typical of English automotive discourse, in this corpus-based study we selected *keywords* and *key terms* as the units of analysis. Identification of keywords is a common practice in corpus linguistic research: according to Brezina (2018, p. 79), “keywords are words that are considerably more frequent in one corpus than

in another corpus; we can therefore say that keywords are words that are typical of the corpus of interest when compared to another corpus". A high frequency in one corpus is not necessarily a characteristic of the keywords. In fact, the most frequent words are not always the most pedagogically useful (Hunston, 2010; O'Keeffe et al., 2007; Szudarski, 2018). Apart from keywords, we also explore so-called key terms, which are noun phrases, or nominalizations, which occur in the EMIC corpus and in each of its subcorpora, with outstanding frequencies compared to the reference corpus. Technically speaking, we identify the keywords and key terms using SketchEngine (Kilgarriff et al., 2014), a custom-designed software for text analysis. More precisely, the simple math metric presented by Kilgarriff (2009), which compares the frequencies of words and nominal phrases in each sub-corpus of the EMIC corpus vs the reference corpus, is used as a keyness metric to identify salient words and terms.⁷ The procedure of keyword extraction was conducted using the SketchEngine software (Kilgarriff et al., 2014). As a reference corpus, we used enTenTen20 (Jakubiček et al., 2013), an all-purpose corpus comprising the texts collected from the Internet covering a multitude of genres and text types that includes more than 43 billion tokens and over 36 billion words types. enTenTen20 was selected because it has broad coverage of text types, genres, registers and varieties of English, and it is of considerable size, which makes it a suitable benchmark for comparisons against specialized text types. Initially, we focus on the top-100 keywords and key terms by keyness with a minimum frequency of 10 occurrences in each subcorpus of the EMIC. Next, we manually filter out proper names and function words.

2.3. Analytical procedures and study stages

The study presented in this paper consists of two separate stages. The first one, which is largely descriptive, focuses on the identification and description of salient keywords and key terms. More precisely, the keywords and key terms were assigned to semantic categories, based on qualitative concordance-based manual analysis of meanings conveyed in particular contexts of specialized language use as recorded in each sub-corpus of the EMIC corpus. This allowed us to describe the usage and semantics of identified items in the English automotive discourse. To that end, we provided examples of use of selected language items in their co-text, arbitrarily selected primarily on the basis of their keyness value, frequency of use and range of language patterns found in the various text types.

In the second stage, which focuses on pedagogical application of descriptive findings, we asked the ESP teachers teaching English automotive language to car mechanics (i.e., the EFL students with A1/B2 language competence levels according to the CEFR) in a selected secondary vocational school in Poland to decide whether each keyword and key term is worth teaching, and if it meets the students' needs, broadly following the approach used by Simpson-Vlach and Ellis (2010). More precisely, three ESP instructors working in one such school in south-western Poland were asked to rate all the keywords and key terms, that is, to decide whether they are worth teaching, and whether they meet the needs of secondary vocational school students trained as car mechanics. This part of the study consisted of the following steps. First, we created the spreadsheets with keywords and key terms (in rows), which included the specific questions (in columns) to be answered (Yes=1, No=0):

1. Is this keyword/key term worth teaching?
2. Does it meet the students' needs?

Before providing the raters with the spreadsheets, we contacted them online to explain the goals of the study and the instructions on how to fill in the prepared forms (Appendix 19). Then, the raters received the online spreadsheets and were asked to answer the aforementioned questions and send back the completed files. We assumed that each keyword and key term should have at least two positive answers, i.e. (1) for both questions, to be considered pedagogically useful. In the last step, the reliability of the raters' assessment of pedagogical worthiness of keywords and key terms was further verified in terms of inter-rater agreement. Reliability is an important aspect of any analytical study involving data classification by annotators, as it ensures that the data is independent of the measuring circumstances (Krippendorff, 2004a, 2004b). We used ReCal3 inter-rater reliability calculator (Freelon, 2010) to measure the percent agreement that does not consider the chance factor, followed by applying Krippendorff's α (alpha) statistics for nominal variables (Yes/No), which is used to assess the reliability between the multiple raters (we had three raters, henceforth R1+R2+R3) taking into consideration their agreement by chance.⁸ The results were interpreted in accordance with Krippendorff's (2004a, 2004b) recommendations: $\alpha = 1.000$ perfect agreement; $\alpha \geq .667$ – acceptable agreement, and $\alpha < .667$ lack of agreement. Krippendorff's α was chosen because it is suitable for small samples of data, it can handle three raters (while Cohen's kappa is limited to two), it accounts for chance agreement, and it is generally preferred for subjective coding, text interpretation or linguistic data coding (Brezina 2018).

2.4. Study goals and research questions

As mentioned earlier, this corpus-based study consists of a descriptive (Section 3: Results: keywords and Section 4 Results: key terms) and applied component (Section 5: Results: selection of pedagogically-useful keywords and key terms). We assume that the analysis of the keywords and key terms found in each sub-corpus of the EMIC corpus will provide a description of salient vocabulary and phraseologies typical of English automotive discourse. Also, we assume that the study will provide a set of pedagogically valuable data that may be later used while developing ESP courses for secondary school students specializing in car mechanics. Hence, the study aims to provide answers to the following research questions: 1) Which keywords and key terms are the most salient in English automotive discourse (i.e., in each sub-corpus of the EMIC corpus); 2) What meanings are conveyed by these keywords and key terms in each sub-corpus of the EMIC corpus; 3) Which keywords and key terms are the most pedagogically useful for teaching ESP to vocational school EFL students with A2/B1 language competence level according to the CEFR.

3. Results: keywords

3.1. Keywords in car reviews (CRs)

After the filtering procedure, we identified 54 keywords in the CRs (Appendix 1), most of them nouns and adjectives. Most nouns in the CRs refer to electronic equipment (“infotainment”, “touchscreen”), parts of a car (“gearbox”, “damper”) or a car’s characteristics (“diesel”, “hatch”). Given a large number of acronyms among keywords, e.g., “AWD” (‘All Wheel Drive’), “SUV” (‘Sport Utility Vehicle’) or “CVT” (‘Continuously Variable Transmission’), we can assume that target readers already know the abbreviated forms, and there is no need to clarify the acronyms’ meanings. Most keywords are directly connected to the construction of a car (Type of car, Engine, Driving and steering system, Part of a car), which corresponds with the thematic scope of CRs. Also, the descriptors are the most frequently used keywords, e.g., “sporty”, “facelifted”, “roomy” or “comfy”. The keywords were classified into seven semantic-functional categories: Type of a car, Technology, Engine, Driving and steering system, Part of a car, Descriptor and Other (Appendix 5).

An adjective “supple” (12 occurrences in CRs) is used to describe a driving experience whereby it suggests that a car has a soft suspension and can easily

adapt to various surfaces or roads. This keyword has a tendency to be used in binomials with another adjective “(adjective) and supple” or is frequently modified by an adverb (e.g. “decently supple”, “remarkably supple”). Another keyword worth closer examination is “roomy” (12 occurrences in CRs), which is used to describe how spacious a car is. This keyword is in comparative constructions (e.g., “(not) as roomy as”), where many cars are compared, or in comparative and superlative forms, as in, e.g. “(...) As roomy as the Ibiza, which means plenty leg, head and (...)” (CR_25); “(...) It’s about the roomiest small crossover (...)” (CR_32).

The keyword “supermini” (32 occurrences in CRs), which is used to describe a small car designed to be comfortable and easy to drive, is modified by adjectives and nouns referring to aesthetic impressions (e.g., “comparable”, “modern”, “mainstream”), car trims or versions (e.g., “step-up”, “sporty”) or names of car brands.

3.2. Keywords in car brochures (CBs)

The analysis revealed 56 keywords in the CBs, (Appendix 2). Among them, the largest group are acronyms and abbreviations (14 items, e.g., “DRCC”, “AWD”, “VTEC” or “nav”), which refer to technology (e.g., “nav” is short for navigation; “DRCC” for ‘Dynamic Radar Cruise Control’), type of engine (e.g., “DCI” for ‘Diesel Common Rail Injection’, “VTEC” for ‘Variable Valve Timing and Lift Electronic Control’). Their high frequency can be explained by the fact that CBs are written by specialists, and they are read by potential car buyers who already have vast knowledge of cars, and they need no clarification.

The analysis of contexts of use of the keywords enabled us to categorize them into eight semantic groups (Appendix 6) related to the construction of a car (Engine, Driving and steering system, Part of a car, Car design and finish, Descriptor) as well as to technologies (Technology) and relevant regulations (Regulations), and the largest number of keywords in the CBs refer to technology. We can assume that the car manufacturers are moving toward more technologically advanced cars, hence the outstanding frequency of the keywords in this category. It is also important to inform potential buyers whether a car complies with applicable legal requirements, hence the acronyms “WLTP” (‘World Harmonised Light Vehicle Test Procedure’), “NEDC” (‘New European Driving Cycle’), “Eu6” (refers to the latest EU emission standard), and “VED” (‘Vehicle Excise Duty, which is a tax in the UK’).

The most frequent keyword in CBs is “rear” (1,691 occurrences). In most cases (981 times), it is used as an adjectival modifier of nouns referring to parts of a car, e.g., “seat”, “bumper”, “parking sensor”, “passenger”, “light” or “window”.

Many interesting language patterns were revealed in the analysis of the keyword “airbag” (216 occurrences), which modifies such nouns as “curtain”, “side”, “passenger” or “seat-mounted”, referring to a location or type of airbag, or deployment, deactivation or system, referring to specific features of the safety system. It also occurs in the object or subject position with respect to verbs such as “inflate”, “deploy”, “activate”, which also describe the functions of an airbag, e.g. “airbag deployment”, “airbag deactivation”, “curtain airbag”, “knee airbag”. Similar patterns were observed with the keyword “seatbelt” (81 occurrences in CBs), which co-occurs with the verbs “wear”, “fasten”, “unfasten”, “ensure” or “help”, describing actions performed by the passengers using seatbelts or their safety advantages.

3.3. Keywords in scripts from motoring TV programme Top Gear (TGSs)

After the application of the cut-off criteria, we obtained the list of 48 keywords in TGSs (Appendix 3). They include many expressions typical of spoken language, e.g., “whoa”, “aah”, “yeah” or “cos” (classified into a separate group of Interjections), as the speakers often express surprise, hesitation, approval, joy or attention when talking about cars in the TV programme. Also, programme hosts often use interjections to catch viewers’ attention and keep their interest by using entertaining language and exclamations.

Based on the concordance analysis, the keywords in TGSs were assigned to seven semantic categories, namely Type of a car, Engine, Driving and steering system, Part of a car, Interjections, Road type and its attributes and Other (Appendix 7). The keywords from the category Road type and its attributes and Types of a car reflect the script of the programme, where either the hosts or invited guests drive a car on a racetrack or unusual terrain, e.g., an unpaved surface, sand, riverbeds or other natural terrain. Hence, the analysis revealed such keywords as “lap”, “racetrack”, or “off-road”. What is more, the keywords from the category Type of a car (e.g., “tuk-tuk”, “track-tor” or “lorry”) are relatively unusual compared to those found in the same

category in CRs and CBs, which may be explained by the fact that the challenges take place not only in the uncommon areas, but also in the unusual vehicles. However, as this study focuses on personal vehicles, the keywords “tuk-tuk”, “lorry” and “track-tor” are not explored further. Also, we ignored interjections in our closer qualitative analysis because they are a characteristic of informal conversations in general rather than the specific discussions between, for example, a car mechanic and a car owner.

The top keyword in TGSs is “supercar” (138 occurrences), which typically refers to an expensive, fast and luxurious sports car. It is often modified by adjectives which provide a more detailed description, e.g., “budget”, “convertible”, “hybrid”, “bland”, “grand”, or “super-expensive”.

Another salient keyword, with the highest frequency in TGSs (387), is “brake” used 247 as a noun and 147 times as a verb. In the analysis, we focused on its nominal use referring to a device that a driver uses to slow down or stop a vehicle. Therefore, the keyword “brake” is modified by adjectives providing a more detailed description of materials used to produce brakes (e.g., “carbon”, “ceramic”) or their size (e.g., “massive”). Also, the keyword “brake” modifies several nouns and is used in specialized terms referring to more specific parts of brakes, e.g., “brake discs”, “brake rotors”, “brake cable”, “brake fluid”.

The keyword “brake” is often used with verbs, namely “touch”, “apply”, “hit”, or “depress”, which specify the actions a driver performs while using the brakes. In a similar vein, we noticed some useful language patterns formed with the keyword “tyre” (247 occurrences), which is frequently modified by such nouns or adjectives as “rear”, “skinny”, “winter”, “front”, “super-skinny” or “ungrippy”. Thanks to these modifiers, the audience obtains specific information about tyre types and their features. Another salient keyword, this time from the Engine category, is “horsepower” (215 occurrences in TGSs), which is a unit of measurement of power in an automobile engine, and it is often used with verbs. We can divide those verbs into two categories, namely those that indicate how powerful an engine is (e.g., “produce”, “develop”, “make”, “give”, “provide”), and those that suggest how much power a driver can attain (e.g., “use”, “deploy” or “add”).

3.4. Keywords in car owner’s manuals (COMs)

In the case of COMs, 80 keywords (Appendix 4) were subjected to detailed concordance-based analysis. As COMs subcorpus consists of the fragments

of car user's manuals devoted to safety, it is no surprise that the top keywords reflected this theme. They refer to the standard vehicle equipment, for instance, "seatbelt", "airbag" or "pretensioner" as well as to optional safety equipment, e.g., "isofix" (which stands for 'International Standard Organisation Fix') or "OWS" ('Occupant Weight Sensor'). We also identified acronyms describing state-of-the-art technologies enhancing passengers' safety, namely "SRS" ('Supplemental Restraint System'), "ALR/ELR" ('Automatic/Emergency Locking Retractor') or "CRS" ('Child Restraint System'). We also found keywords describing collision types, e.g., "rollover" or "side-impact". The keywords in COMs were divided into eight semantic categories, namely Technology, Safety devices, Car parts, Location, Descriptor, Accident (Appendix 8).

The category with the largest number of keywords is Safety devices: these inform car users about different types of safety equipment in a vehicle (e.g., "restraint", "protection", "warning"), and some of them provide a detailed description of each part of a seatbelt (e.g., "webbing", "pretensioner", "limiter", "retractor" or "latch"). The keyword with the highest frequency in COMs is "seat" (5,971 occurrences). It is mainly used as a noun (5,558 instances) and indicates a location of a passenger, e.g., "front", "rear", "forward-facing" or "rear-facing". Also, it is often modified by nouns or adjectives that specify a type of seat, e.g. "child seat", "passenger seat", "booster seat". The keyword "seat" also functions as an object of verbs such as "move", "adjust", "install", "use", "occupy", "slide", "attach", "fit", which specify actions performed by a passenger.

Another keyword, "seatbelt" (172 occurrences), is often used as an object of verbs describing the actions performed by a passenger, as in CBs. However, we detected different verbs from those in CBs, namely "buckle", "unbuckle", "retract", "swing", "replace" and "release". What is more, this keyword frequently modifies the keywords from the category Safety devices, e.g., "seatbelt retractor" (a mechanism holding a webbing material and allowing to feed it out or lock it in case of strong pull), "seatbelt extension" (additional part of webbing added temporarily if original seatbelts are not long enough), "seatbelt pretensioner" (a mechanism that tightens up webbing in the event of collision), "seatbelt tongue" (a metal part of a seatbelt that a passenger fits into the buckle), "seatbelt adjuster" (it helps a passenger to regulate webbing so that it fits a body).

4. Results: key terms

4.1. Key terms in car reviews (CRs)

After filtering out proper names, most of them referring to car brands or models, and combining phrases that were part of longer nominalizations (e.g., “media system” and “touchscreen media” make up a longer compound “touchscreen media system”), we identified 71 key terms in CRs (Appendix 9). Then, we classified the key terms into nine semantic categories (Appendix 13) based on their meaning in co-text and context, which were found to be parallel to those for keywords in CRs, with the exception of two additional, more fine-grained categories, namely Powertrain and Gearbox. The categories are as follows: Type of a car, Technology, Engine, Driving and steering system, Part of a car, Descriptor, Gearbox, Powertrain and Other.

The analysis revealed that most categories refer to the construction of a car (Engine, Part of a car, Gearbox, Powertrain). Also, in the Descriptor category, we find 14 items that provide detailed characteristics of particular car features or parts. For example, the noun phrase “trim level” (14 occurrences in CRs, category Part of a car) specifies a vehicle’s level of equipment, where the vehicle equipped with the least features is called a “base model” (key term from the category Descriptor, with 6 occurrences in CRs), and it is modified by words such as “trim”, “lower”, “top”, “four”, “top-of-the-range”, “higher”, “entry-level”, e.g.: “(...) It would make more sense if these phone mirroring functions also appeared on the lower trim levels which don’t have built-in navigation. (...)” (CR_23)

The most numerous category is Technology (17 key terms), where nominalizations refer to entertainment systems (e.g., “infotainment system/screen”, “phone mirror” or “touchscreen media system”) and safety features (e.g., “blind-spot warning”, “adaptive cruise” or “parking sensor”). This can be explained by technological developments enhancing passengers’ comfort and safety and CRs readers’ interest in such state-of-the-art solutions. For example, the term “cruise control” (11 occurrences) is either used as an ‘automatic control that adjusts engine output and selects gear in automatic gear change system, to maintain a constant pre-selected speed’ (Goodsell, 2016, Chapter C) or in longer lexical strings, such as “adaptive cruise control” or “radar cruise control”, which refer to ‘a system that automatically keeps a safe space from the vehicle ahead by a radar and control of the accelerator and brake’ (Lane, 2002, section 1).

The top-key term in CRs is “hot hatch” (21 occurrences), which is a short form of hot hatchback, ‘a powerful version of a small hatchback car’ (Lane, 2002, section 8). It is frequently modified by adjectives such as “sporty”, “amusing”, “classic” or “first-ever”, as in: “(...) Choose the standard car and you’ll enjoy an amusing hot hatch with a twist (...)” (CR_34)

4.2. Key terms in car brochures (CBs)

As in CRs, we applied cut-off criteria and excluded proper names and shorter noun phrases included in longer ones (e.g., “hand drive” and “drive vehicle” making up a longer phrase “left/right hand drive vehicle”). Then we identified 65 key terms in CBs (Appendix 10), which were further classified into seven semantic categories, namely Technology, Driving and steering system, Part of a car, Descriptor, Car design and finish, Exhaust emission and Other (Appendix 14).

Many key terms in CBs refer to the construction of a car (Part of a car, Descriptor, Car design and finish), driving (Driving and steering system and Technology) and environmental issues (Exhaust emission), which is a new category emphasizing growing interest of car users and buyers in environmental legal regulations, carbon footprint and resulting costs associated with car maintenance. For example, two noun phrases from the latter category, which are “fuel consumption” (222 occurrences) and “CO2 figures” (108 occurrences), often occur together in CBs, e.g., “(...) Further information about the test used to establish fuel consumption and CO2 figures can be found at (...)” (CB_21).

In a similar vein, the noun phrase “climate control” (99 occurrences) is often used in the context of car features designed to control the inside temperature in a vehicle or the level of air pollution and humidity, e.g., “air conditioning”, “air quality regulator”, “demist function” or “humidity sensor”. It is often followed by modifiers such as “dual-zone”, “tri-zone”, “manual and automatic”, as in: “(...) Automatic dual zone climate control with pollen filter (...)” (CB_19).

Much as in CRs, the most numerous category is Technology (18 items), with several overlapping key terms, e.g., “climate control”, “cruise control”, “centre console”, “parking sensor” or “infotainment system”. We also identified several other key terms used in longer strings, e.g., adjectives “rear” and “front” modify “parking sensor”; “warning” and “alert” follow “lane departure”; “blind spot” is frequently followed by “detection”, “warning”

and “information”; “keyless entry” is followed by “system”, while “rear-view” by “display”; the adjective “active” modifies such key terms as “info display” and “park assist”, as in: “(...) The front and rear parking sensors guide you both in and out of any parking position. (...)” (CB_ 14).

The key term with an outstanding frequency in CBs is “alloy wheel” (764 occurrences) from the category Car design and finish. The alloy wheels are ‘road wheels made from an aluminium-magnesium alloy, much lighter than steel wheels’ (Lane, 2002, section 1). In contrast, the key term “steel wheel” appears in CBs only 46 times. Although this difference results from specific car models described in CBs, we may assume that alloy wheels are preferred by car manufacturers and car users because they are lightweight and robust and are often considered to be stylish. The noun phrase “alloy wheel” is modified by descriptors specifying its characteristics, e.g., “machined-finish”, “polished”, “chrome-finished”, “metallic”, “bi-coloured”, “sport”, “white”, “black”, “silver”, “unique”, as in: “(...) Metal machined-finish alloy wheels (...)” (CB_10).

The key term “lumbar support” (55 occurrences in CBs) provides additional information about the seats that include special support for lower backs, and this way they enhance passengers’ comfort. Also, lumbar support is frequently modified by a type of adjustment system, e.g., “manual”, “electrical” or “power-adjustable”, as in: “(...) Driver’s seat with manual lumbar support (...)” (CB_13).

4.3. Key terms in scripts from the motoring TV programme Top Gear (TGSs)

Having filtered out proper names, including names of hosts and guests of the TV programme, car brands and models and phrases typical of general or spoken English (e.g., “good night” or “bloody hell”), we identified 57 key terms in TGSs (Appendix 11), which were later assigned to the eight semantic categories that reflect the content of the scripts from the motoring TV programme “Top Gear”, which are as follows: Type of a car, Engine, Driving and steering system, Part of a car, Descriptor, Road type and its attributes, Racing and Other (Appendix 15).

As compared to keywords in TGSs, there is no category Interjections, but two additional categories were introduced, namely Descriptor and Racing, which are the most numerous. The noun phrases in the Descriptor category can be divided into several sub-categories, namely referring to the car’s

appearance (“little car”, “black cab”); expressing a personal opinion about a car (“priced car”, “great car”, “favourite car”, “good car”, “fast car”); referring to a car’s suitability (“road car”, “city car”); and specificity (“muscle car”, “classic car”, “American car”), as in: “(...) But now, it is time to introduce to you all a brand-new reasonably priced car (...)” (TGS_42)

A peculiar characteristic of TGSs is the category Racing (e.g., “finish line”, “final track”, “race track”, “drag race”) because the programme’s purpose is to present car-connected issues in an entertaining manner. For example, the hosts invite guests and racing drivers to participate in a race as a challenge. In that category, the compound noun “finish line” (32 occurrences) is outstanding, and it often co-occurs with verbs such as “reach”, “push over”, “cross”, “race” or “gallop”, as in: “(...) You’re racing towards the finish line, neck and neck, you want to cross the line first. (...)” (TGS_35).

4.4. Key terms in car owner’s manuals (COMs)

In the COMs, we noticed that several key terms (e.g., “air bag”) were fragments of longer strings, such as “air bag deactivation” and “indicator light”, which make up even longer lexical items (“air bag deactivation indicator light”), and that is why we decided to use the longest sequences of words in the analysis.⁹ This resulted in the final list of 55 key terms in COMs (Appendix 12), which were further classified into six semantic categories: Technology, Safety devices, Car parts, Descriptor, Accident and Other (Appendix 16).

The most numerous category in the COMs is Safety devices, which provide specific information on various safety solutions, namely airbags (e.g., “air bag”, “front airbag”, “passenger air bag”, “side curtain”, “airbag system”); seatbelts (e.g., “safety belt”, “seat belt pretensioner”, “belt extender”, “automatic locking mode”, “lap belt”, “front seat belt”, “belt buckle”, “top tether”, “tether anchor”); and restraints (e.g., “child restraint system”, “head restraint”, “supplemental restraint”).

In the case of seatbelts, we noticed that two nominalizations, “shoulder belt” (177 occurrences) and “lap belt” (55 occurrences), are often used in recommendations or explanations as to how to position the seatbelt’s webbing to be safe and comfortable. That is why they co-occur with adverbs of frequency (“always” or “never”), modal verbs (“must”, “should” and “could”), which describe recommended actions followed or preceded by if-clauses specifying risks connected with inappropriately

positioned webbing, as in “(...) Never route the lap belt across your abdomen. (...)” (COM₁₁).

We also noticed that the key term “top tether” (164 occurrences), a part of a child restraint system used with a child seat, tends to co-occur with select lexical items (e.g., “strap”, “anchor”, “belt” or “system”) making up more extended specialized terms, which refer to particular parts of the top tether device: the nominalization “top tether system” indicates a whole mechanism, as in: “(...) Make sure that you do not interfere with the correct routing of top tether belt (...)” (COM₁₁).

A top key term in COMs is “child restraint” (1,095 occurrences), which refers to ‘a device for preventing a load from moving, including ropes, straps, or chains’ (Lane, 2002, section 18), and which is often used in a longer nominalization “child restraint system”. It is frequently used with the verbs “install” or “secure”, and it is modified by words specifying direction of positioning that device, e.g., “rearward-facing”, “rear-facing” and “forward-facing”, as in: “(...) Follow these steps to install a rear-facing child restraint using the LATCH system (...)” (COM₉).

The shorter key term, “restraint system” (670 occurrences), is modified by adjectives describing its characteristics or condition, e.g., “malfunctioning”, “suitable”, “additional” or “supplemental”, as in: “(...) Risk of injury or death when additional restraint systems are not used for persons with a smaller build (...)” (COM₁₁).

In the category Technology in COMs, we found that nominalizations such as “indicator light”, “warning light” and “warning lamp” are used with a small set of verbs (e.g., “flash”, “illuminate”, “light up”, “turn on”) referring to operation of these signals in the event of car malfunctioning, as in: “(...) The indicator light lights up when a child is properly seated in a child restraint system (...)” (COM₁₀).

5. Results: selection of pedagogically-useful keywords and key terms

In this applied part of the study, we aim to select the most pedagogically-useful keywords and key terms from the entire EMIC corpus to be used in training future car mechanics in secondary vocational schools. Based on the results of assessment conducted by ESP teachers, we obtained a list of 89

keywords (Appendix 17) and 125 key terms (Appendix 18), which have pedagogical value in teaching secondary vocational school students, future car mechanics, with A2/B1 EFL competence in Poland.

As for the pedagogically useful keywords, we can see that the annotators did not mark positively the interjections in TGSc (e.g., “whoa”, “cos”, “yeah”, “um”) that do not convey specialized meanings relevant to the motor industry. Also, the final list includes a relatively small number of acronyms and abbreviations. It appears that the ESP teachers prefer to teach the full terms when given a choice between a full form and an acronym, for instance, all-wheel drive and AWD. Based on our perfunctory inspection of a number of acronyms in the National Corpus of Polish (NKJP), we assume that such a preference results from the fact that the English acronyms of technical and specialised terms are also commonly used in Polish. Also, many keywords referring to safety were ranked as not pedagogically useful as they refer to very specific, individual car components, e.g., “pretensioner”, “retractor”, “tensioner” or “strap”. Given a limited number of 30 teaching hours dedicated to specialized languages, including ESP, in Polish secondary vocational schools, the teachers have to weigh the value of teaching materials and may prefer to teach more general words, such as seatbelt, instead of focusing on individual components of seatbelts. Furthermore, such specific vocabulary may be demanded from more advanced students at B2 or C1 levels of competence but not at A2 or B1 levels. Overall, the largest number of pedagogically useful keywords refer to parts of a car, engine, driving and steering system, and descriptors.

Furthermore, we observed that the annotators decided to positively code more than half of all key terms (55%), which is a considerable difference compared to the results from keywords, where only 42% were considered as pedagogically useful. It may be due to the more specialized meanings of the key terms, which resulted in a more accurate comprehension of the terms on the part of the annotators, e.g., the keyword “supple” was not marked as worth teaching, while the key term “supple ride” was rated as a pedagogically useful phrase related to driving experience. In the same vein, the ESP teachers marked positively the key terms such as “navigation system”, “optional equipment” or “child restraint system”, and negatively the keywords such as “nav”, “optional”, “child-restraint” or “CRS”. Also, we identified numerous instances of car types with their descriptors marked as worth teaching, e.g., “sports car”, “small car”, “standard car”, “rally car”, “classic car”, “fast car” or “old car”. As these key terms might reveal

personal opinions about a specific vehicle, as well as information about a car's suitability or specificity, they may be used in simple communications in a workplace. Overall, most of the pedagogically useful key terms represent the categories of Parts of cars, Engine, Driving and Steering System, and – in contrast to keywords – we noted a high number of pedagogically useful key terms in the categories of Technology (e.g., “adaptive cruise control”, “keyless entry system”) and Safety (e.g., “airbag system”, “belt buckle”).

In the last stage, the reliability of the annotators' assessment of pedagogical worthiness of keywords and key terms was further verified in terms of inter-rater agreement. More precisely, the inter-rater agreement for keywords in percent was 68.39% for Question 1 and 67.76% for Question 2, indicating a substantial agreement, and the corresponding figures for Krippendorff's α were 0.36 and 0.35, suggesting low agreement between the annotators. As for the key terms, the percent agreement for Question 1 and Question 2 was 54.48% and 51.84 % respectively, suggesting a moderate agreement, while the corresponding Krippendorff's α scores were 0.04 and 0.003, suggesting almost no agreement between the annotators.

6. Discussion and conclusions

This descriptive and applied corpus-based study provides a comprehensive description of lexical variation across four text types found in English language used in various automotive contexts, namely car reviews, car brochures, transcripts of a motoring TV programme and fragments of car user's manual devoted to safety. The aforementioned text types feature a high degree of linguistic variation in terms of the frequency and use of keywords and key terms, which we adopted as the units of analysis. The differences result from situational and functional characteristics of these text types, which we explained in the qualitative analysis. Also, we made an attempt at identifying the most pedagogically useful keywords and key terms that can be used in vocational education.

According to the relevant core curriculum, car mechanics trained in secondary vocational schools in Poland need to learn ESP to an extent allowing them to communicate in routine professional contexts. Therefore, they should know basic vocabulary in order to be able to perform a range of actions related to the workplace and its equipment, the technologies used in the profession, the documentation and the services provided. What is more,

they are expected to understand and create simple oral expressions and written texts (e.g., communications, instructions, presentations, brochures, manuals, guides and professional documents) to be able to carry out professional tasks. Hence, the students need to acquire vocabulary and phraseology referring to car parts, types of engines or steering systems, and such keywords and key terms were selected in our study by the ESP teachers as worth teaching, irrespective of low level of inter-rater agreement. Brezina (2018, p. 89) noted that “there is no magic number beyond which (raw) agreement becomes universally acceptable”, which means that our interpretation considers the study’s context: the selected vocabulary and phraseology should meet the criteria required by the core curriculum and learning outcomes for the ESP classes in the secondary vocational schools in Poland. Therefore, we can plausibly argue that our findings may be successfully adopted to create the ESP course syllabi, teaching materials, developing testing and assessment. Moreover, these lexical items can be used in spoken and written communication at workplace with colleagues, customers, and employers, as well as in situations such as reporting car malfunctions, comprehending car owner’s manuals, broadening one’s knowledge and keeping up-to-date with the innovations from the field of the motor industry, among others.

The study findings provided answers to our research questions. As regards R1, we identified salient keywords and key terms typical of each text type under scrutiny. As for R2, we classified the keywords and key terms into a number of semantic categories in each text type, e.g., type of a car, technology, engine, driving and steering system, car parts, safety devices, or descriptors of a car. What is more, we noticed that the keywords and key terms are used to convey various meanings related to key topics in the analyzed text types, e.g., they provide information of a car type (e.g., “supercar”, “hatchback”, “sports car”); they refer to a particular part of a car (e.g., “chassis”, “gearbox”, “brake light”); they describe type of an engine (e.g., “electric motor”, “petrol engine”, “valvematic”); they refer to the technology used in a vehicle (e.g., “infotainment”, “keyless entry”, “nav”); they describe driving and steering system (e.g., “AWD”, “torque”, “top speed”); they express personal opinions about a vehicle (e.g., “comfy”, “sporty”, “hot hatch”); they report on safety features (e.g., “child-restraint”, “airbag”, “seatbelt”). Finally, as regards R3, we also revealed the lists of 89 keywords and 125 key terms which were found to be pedagogically useful and meet the requirements established in the core curriculum for the ESP

classes at secondary vocational schools in Poland. We argue that the higher number of key terms results from the fact that they are more neatly tied to the specialized domain of language use in comparison to keywords, as many of the latter ones may be also used in general language.

Hence, we can plausibly argue that the pedagogically-useful vocabulary and phraseology should be adopted by instructors at secondary vocational schools in Poland to create the ESP course syllabi, teaching materials, developing testing and assessment. For example, the teachers may design vocabulary exercises (e.g., fill-in-the gaps, word formation or multiple-choice tasks) using keywords and key terms. The lexical items can be also used in speaking tasks and role-playing games (e.g., car description, reporting a malfunction, explaining a piece of technology used in a car). Learners can also be asked to write a text (e.g., a car review, brochure or a short manual) with a given keyword or key terms. We also assume that the keywords and key terms can be applied in other vocational training contexts connected with the automotive sector, e.g., car electro mechanics, mechatronics or engineering.

The study is not devoid of limitations. The size of the study corpus and the criteria for sample selection (we focused on top-100 keywords and key terms, which were subsequently filtered out) could have resulted in information loss. Next, the rather unsatisfactory results of inter-rater agreement may be due to the fact that the ESP teachers use their own limited knowledge of a specialized field and intuition when compiling teaching materials for ESP courses. Consequently, ESP courses may vary among the teachers, as each of them has different opinions on which particular language items should be taught in the first place. This may explain why they might have rated different language structures positively depending on particular skills and learning effects (e.g., workplace safety, repairing a car, car equipment), among others. Perhaps the list of annotators should be extended in the future by including the professionals in the field of motor industry. Employers, experts in the field of motor industry or experienced car mechanics, preferably native speakers of English, could help rate the lists of keywords and key terms to identify those with the highest demand in the car mechanics' workplace. Finally, the findings obtained with the Polish EFL secondary vocational schools students' needs in mind would benefit from additional assessment, e.g., in accordance with the criteria for wordlist evaluation proposed by Nation (2016).

In the future, the study can be expanded in several ways. More text types could be added to the corpus so that it provides more in-depth insights into the English automotive discourse. Such a larger corpus might include, for example, transcripts from oral communication from the workplace, motor industry blogs, transcripts of motor industry podcasts. The additional texts could provide a wider thematic scope of vocabulary and phraseology than the one identified in this study, e.g., pertaining to tools, occupational health and safety. We could enrich the research material by compiling comparable corpora of corresponding automotive text types originally produced in languages other than English, all in order to develop various types of multilingual resources (e.g., electronic glossaries, lexical databases, to name but a few examples).

Furthermore, a comparison of spoken and written registers could also be beneficial for a more comprehensive description of English used in automotive contexts. It could reveal which type of communication is more formulaic and what the differences are between spoken and written specialized discourse (cf. Biber et al. (2004) for investigation of spoken and written registers in English used for academic purposes).

All in all, we believe that the study findings provide a comprehensive and systematic description of the English language used in the motor industry, which has pedagogical potential in the context of vocational education at secondary school level. The study may also provide an excellent starting point for further research on English automotive discourse, which may be conducted using other methods than corpus linguistic ones, which has been the case in this study.

Appendices

Link to Supplementary data (Appendix 1-21): <https://osf.io/fmbs2>

Article history
Received 8 August 2024
Received in revised form 13 March 2025
Accepted 13 May 2025

References

- Benson, S., & Coxhead, A. (2022). Technical single and multiword unit vocabulary in spoken rugby discourse. *English for Specific Purposes*, 66, 111-130. <https://doi.org/10.1016/j.esp.2022.02.001>
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. <https://doi.org/10.1093/applin/25.3.371>
- Boreus, K., & Bergstrom, G. (2017). *Analyzing text and discourse: Eight approaches for the social sciences*. Sage.
- Boukes, M. (2019). Infotainment. In T. P. Vos & F. Hanusch (Eds.), *The International Encyclopedia of Journalism Studies* (Vol. 2) (pp. 1-9). Wiley Blackwell. <https://doi.org/10.1002/9781118841570.iejs0132>
- Brezina, V. (2018). *Statistics in corpus linguistics*. Cambridge University Press.
- Cambridge Dictionary (n.d). Air-bag. In dictionary.cambridge.org. Retrieved October 27, 2024, from <https://dictionary.cambridge.org/dictionary/english/air-bag>
- Chang, J. (2014). The use of general and specialized corpora as reference sources for academic English writing: A case study. *ReCALL*, 26(2), 243-259. <https://doi.org/10.1017/S0958344014000056>
- Collins (n.d.). Airbag. In collinsdictionary.com. Retrieved October 27, 2024, from <https://www.collinsdictionary.com/dictionary/english/airbag>
- Coxhead, A. (2018). *Vocabulary and English for Specific Purposes Research. Quantitative and qualitative perspectives*. Routledge.
- Coxhead, A. (2022). Corpora for teaching and learning vocabulary in ESP. In R. Jablonkai, & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 193-205). Routledge.
- Coxhead, A., & Demecheleer, M. (2018). Investigating the technical vocabulary of plumbing. *English for Specific Purposes*, 51, 84-97. <https://doi.org/10.1016/j.esp.2018.03.006>
- Coxhead, A., Parkinson, J., Mackay, J., & McLaughlin, E. (2020). *English for Vocational purposes language use in trades education*. Routledge.
- Egbert, J., & Biber, D. (2019). Incorporating text dispersion into keyword analyses. *Corpora*, 14(1), 77-104.
- Forey, G. (2004). Workplace texts: do they mean the same for teachers and business people? *English for Specific Purposes*, 23(4), 447-469. <https://doi.org/10.1016/j.esp.2004.01.003>
- Fox, J., & Tigchelaar, M. (2015). Creating an engineering academic formulas list. *The Journal of Teaching English for Specific and Academic Purposes*, 3(2), 295-304.
- Freelon, D. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, 5(1), 20-33.
- Gabrielatos, C. (2018). Keyness analysis: Nature, metrics and techniques. In C. Taylor, & A. Marchi (Eds.), *Corpus approaches to discourse: A critical review* (pp. 225-258). Routledge.
- Ghadessy, M., Henry, A., & Roseberry, R. L. (Eds.). (2001). *Small corpus studies and ELT: Theory and practice*. John Benjamins.
- Gilmore, A., & Millar, N. (2018). The language of civil engineering research articles: A corpus-based approach. *English for Specific Purposes*, 51, 1-17. <https://doi.org/10.1016/j.esp.2018.02.002>
- Goodsell, D. (2016). *Dictionary of Automotive Engineering*. Butterworths.
- Goźdz-Roszkowski, S. (2021). Corpus linguistics in legal discourse. *International Journal for the Semiotics of Law*, 34, 1515-1540. <http://dx.doi.org/10.1007/s11196-021-09860-8>
- Grabowski, Ł. (2015). Keywords and lexical bundles within English pharmaceutical discourse: a corpus-driven description". *English for Specific Purposes*, 38, 23-33. <https://doi.org/10.1016/j.esp.2014.10.004>
- Grzesiak, E. (2021). Career guidance at school and in eTwinning. In D. Tokarz (Ed.), *Vocational education and training with eTwinning. Analyses and inspiration from Polish schools* (pp. 19-26). Foundation for the Development of the Education System
- Hanks, E., Hashimoto, B., & Egbert, J. (2024). The contracts word list: Integral vocabulary for reading and writing English contracts. *English for Specific Purposes*, 25, 37-48. <https://doi.org/10.1016/j.esp.2024.03.002>
- Holmes, J. (2015). Discourse at workplace. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The Handbook of Discourse Analysis* (pp. 880-901). John Wiley & Sons. <https://doi.org/10.1002/9781118584194.ch41>
- Hunston, S. (2010). *Corpora in Applied Linguistics*.

Cambridge University Press.

Jablonkai, R. (2022). Building corpora for ELT. In R. Jablonkai, & E. Csomay (Eds.), *The Routledge handbook of corpora and English language teaching and learning* (pp. 460-477). Routledge.

Jakubiček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. In *7th International Corpus Linguistics Conference CL* (pp. 125-127).

Kilgarriff, A. (2009). Simple maths for keywords. In M. Mahlberg, V. González-Díaz, & C. Smith (Eds.), *Proceedings of the Corpus Linguistics Conference* (pp. 41-55). University of Liverpool.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography*, 1, 7-36. <https://doi.org/10.1007/s40607-014-0009-9>

Koester, A. (2004). *The Language of work*. Routledge.

Koester, A. (2006). *Investigating workplace discourse*. Routledge.

Koester, A. (2010). Building small specialized corpora. In A. O'Keefe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 66-79). Routledge.

Krippendorff, K. (2004a). *Content analysis. An introduction to its methodology*. Sage.

Krippendorff, K. (2004b). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, 30(3), 411-433.

Lane, K. (2002). *Automotive A-Z: Lane's complete dictionary of automotive terms*. Veloce Publishing.

Lonsmann, D., & Kraft, K. (2017). Language in blue-collar workplaces. In B. Vine (Ed.), *The Routledge handbook of language in the workplace* (pp. 138-149). Routledge.

Martinez, R., & Schmitt, N. (2012). A phrasal expressions list. *Applied Linguistics*, 33(1), 299-320. <https://doi.org/10.1093/applin/ams010>

Merriam-Webster (n.d.). Airbag. In Merriam-Webster.com dictionary. Retrieved October 27, 2024, from <https://www.merriam-webster.com/dictionary/airbag>

Ministry of National Education of Poland (2018). Mechanik pojazdów samochodowych. *Dziennik Ustaw Rzeczypospolitej Polskiej*. Retrieved February 15, 2024, from <https://www.ore.edu.pl/wp-content/uploads/2020/03/mechanik-pojazdow-samochodowych.pdf>

Ministry of National Education of Poland (2019). Rozporządzenie Ministra Edukacji Narodowej zmieniające rozporządzenie w sprawie podstawy programowej wychowania przedszkolnego oraz podstawy programowej kształcenia ogólnego dla szkoły podstawowej, w tym dla uczniów z niepełnosprawnością intelektualną w stopniu umiarkowanym lub znacznym, kształcenia ogólnego dla branżowej szkoły I stopnia, kształcenia ogólnego dla szkoły specjalnej przysposabiającej do pracy oraz kształcenia ogólnego dla szkoły policealnej. *Dziennik Ustaw Rzeczypospolitej Polskiej*. Retrieved February 9, 2024, from <https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20180001679/O/D20181679.pdf>

Mudraya, O. (2006). Engineering English: A lexical frequency instructional model. *English for Specific Purposes*, 25(2), 235-256. <https://doi.org/10.1016/j.esp.2005.05.002>

Nation, P. (2016). *Making and using word lists for language learning and testing*. John Benjamins.

O'Keefe, A., McCarthy, M., & Carter, R. A. (2007). *From corpus to classroom. Language use and language teaching*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511497650>

Parkinson, J., Coxhead, A., Demecheleer, M., Mackay, J., Matautia, L., McLaughlin, E., & Tu'amoheloa, F. (2017). The language in the Trades education project. Ako Aotearoa.

Scott, M. (2008). Developing WordSmith. *International Journal of English Studies*, 8(1), 95-106.

Simpson-Vlach, R., & Ellis, N. C. (2010). An academic formulas list: New methods in phraseology research. *Applied Linguistics*, 31(4), 487-512. <https://doi.org/10.1093/applin/amp058>

Sowa, M. (2022). The needs of LSP teachers in the preparation and implementation of the didactic process. An attempt at diagnosis. In J. Kic-Drgas, & J. Woźniak (Eds.), *Perspectives on LSP teacher training in Poland* (pp. 28-52). Foundation for the Development of the Education System.

Szudarski, P. (2018). *Corpus linguistics for vocabulary*. Routledge. <https://doi.org/10.4324/9781315107769>

Wang, J., Liang, S.-I., & Ge, G. (2008). Establishment of a medical academic word list. *English for Specific Purposes*, 27(4), 442-458. <https://doi.org/10.1016/j.esp.2008.05.003>

Yang, M. N. (2015). A nursing academic word list. *English for Specific Purposes*, 37, 27-38. <https://doi.org/10.1016/j.esp.2014.05.003>

Justyna Kendik-Gut, PhD, is Lecturer at Foreign Language Centre at the University of Silesia in Katowice, Poland, where she teaches Practical English and ESP. Her research interests include corpus linguistics, teaching English as a foreign language and teaching English for specific purposes. She has published journal articles in *Theory and Practice of Second Language Acquisition* and *Corpus Linguistics Research*.

Łukasz Grabowski, PhD, Prof., is Professor of Linguistics at the Institute of Linguistics, University of Opole, Poland. His main research interests include corpus linguistics, formulaic language, translation studies and computer-assisted methods of text analysis. In 2013, he was a post-doctoral research fellow at the Centre for Advanced Research on English (CARE) at the University of Birmingham (UK). He also worked at the University of Ostrava (Czechia) and, on a number of corpus projects, at the University of Łódź (Poland) and Aston University (UK). He has recently published on semantic prosody, phraseology markers and breakers, as well as on formulaic language in translation and interpreting.

NOTES

¹ Ministerial Decree by the Ministry of National Education of Poland dated 26 July 2018: <https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20180001679/O/D20181679.pdf> (accessed February 9, 2024)

² Description of the qualifications (<https://www.ore.edu.pl/wp-content/uploads/2020/03/mechanik-pojazdow-samochodowych.pdf>) is found in the Regulation of the Minister of National Education of Poland dated 16 May 2019 (item 991), as amended accessed February 15, 2024), pages 2301-2303.

³ The magazine, however, did not mention the publication date or a reviewer's name. The data collection started on 24 February 2018 and finished on 15 March 2018.

⁴ The brochures were collected from 4 February to 30 June 2019.

⁵ The scripts were collected from 16 July to 18 July 2019 from the following website: https://www.springfieldspringfield.co.uk/episode_scripts.php?tv-show=top-gear-2002

⁶ They were retrieved from 4 January 2020 to 28 March 2020.

⁷ Similar to traditional keyword metrics (relying on tests of statistical significance and effect size, cf. Scott 2008; Gabrielatos, 2018), the Simple math metric is also based on comparisons of word frequencies. Recently, an alternative keyword metric based on word distribution (or text dispersion) across texts was developed (Egbert & Biber 2019). In the future, it might be possible to compare the results obtained using these two alternative methods.

⁸ The alpha is defined by an equation: $\alpha = 1 - D_o / D_e$, “where D_o is a measure of the observed disagreement and D_e is a measure of the disagreement that can be expected when chance prevails” (Krippendorf 2004a, p. 222).

⁹ According to Cambridge Dictionary, Merriam-Webster Dictionary and Collins Cobuild Dictionary, the two versions of spelling (air bag and airbag) are acceptable, so we decided to keep them both on the list of key terms.

