Abstract

The aim of this thesis is to examine the linguistic phenomenon of pragmatemes from contrastive and translational perspectives. To achieve this aim, English, French, and Polish conversational pragmatemes found in corpora consisting of various TV series captions are analyzed.

Pragmatemes are language units that often go unnoticed in everyday life. Examples include *Cash or credit?*, asked by a shop assistant, *Happy birthday*, wished on one's birthday, *Hello?*, uttered when picking up the phone, and *Bless you*, uttered after somebody sneezes. These examples, although linguistically different from each other (e.g., *Hello?* is monolexical and the others are polylexical, the meaning of *Cash or credit?* is composed of the meanings of individual words, while the meaning of *Bless you* is not, etc.), have one property in common: they are all typically uttered in a specific situation of communication. Pragmatemes are, therefore, a key element of language fluency, as using different utterances in the situations discussed above may result in sounding unnatural (in the best case scenario) or in a misunderstanding (in the worst). For instance, if one says *On your health* (a literal translation of Polish *Na zdrowie*) instead of *Bless you*, they will not be understood, as such a phrase may rather resemble the one used for raising a toast (*To your health*).

In literature, the phenomenon of pragmatemes (or similar notions) has been described under various names (e.g., Coulmas's 1979 'routine formulas', Kecskés's 2000 'situation-bound utterances', and Kauffer's 2019 'stereotyped language acts', among others). Furthermore, across different studies on the topic, varying definitions have been employed, often with mutually excluding traits (e.g., for Blanco and Mejri [2018], a 'prototypical pragmateme' is compositional, but Kauffer [2019] defines his 'stereotyped language acts' as idiomatic units). These definitions also frequently apply to one language (i.e., the language of the study) but lack accuracy when applied to equivalent units in another language. For example, the expression commonly used for raising a toast, *Cheers*, consists of one lexeme in English but two in Polish (*Na zdrowie*). This means that the utterance used in the equivalent situation in both languages can be considered a pragmateme in Polish but not in English if the composition of two or more lexemes is regarded as a defining factor (such as for Blanco

and Mejri [2018] in regards to their 'prototypical pragmateme'). Therefore, this study's first research question aims to propose a universally applicable definition of pragmatemes. Following an extensive literature review, the proposed definition states that pragmatemes are fixed language units predictably used in typical, repeatable, and specific situations of communication, and that they independently trigger a mental representation of said situations. Additionally, three types of pragmatemes are suggested: a plain pragmateme is a unit whose meaning is based on the individual meanings of its components (i.e., is compositional), a loaded pragmateme is idiomatic, i.e., its meaning is not derived from the individual meanings of its components, and a charged pragmateme is an ambiguous unit, either compositional or idiomatic, the interpretation of which depends on the context.

With the notion of a pragmateme explained, it is now appropriate to discuss other research questions of this study. The second research question concerns the linguistic characteristics of pragmatemes. Previous linguistic research on the topic has primarily approached it from a lexicographic perspective (e.g., Krzyżanowska, Grossmann, and Kwapisz-Osadnik 2021), without delving into extensive investigations into the linguistic traits of pragmatemes across different languages. Particularly, there is a dearth of studies investigating English, French, and Polish units together. Therefore, this study aims to fill this gap by analyzing conversational pragmatemes from the perspective of their linguistic characteristics. The goal is to verify whether they are as irregular as they may appear and to examine whether there are any linguistic similarities among pragmatemes regardless of a language. For this purpose, pragmatemes from three different language families were examined: English, a Germanic language, French, a Romance language, and Polish, a Slavic language.

Given that "in pragmatics research frequency-based research is relatively rare" (Wood 2015: 94), the method of starting with preidentified formulas was applied. Therefore, in the initial stage of the research, pragmatemes were extracted from 45 different reliable sources such as language dictionaries and phrasebooks. The preidentified pragmatemes were then searched for in the corpora of captions of 11 English, 20 French, and 10 Polish TV series (and four movies in terms of the Polish corpus) available on Netflix, which accounts for 770,836 tokens in the English corpus, 664,508 tokens in the French corpus, and 138,516 tokens in the Polish corpus. In total,

290 English, 186 French, and 106 Polish pragmatemes were found and examined based on the following characteristics: frequency in the corpora, complexity (i.e., the number of words), variantivity, imperativeness, presence of verbless forms, presence of question forms, presence of elliptical forms, presence of deictic expressions, speech acts, and pragmateme types. The analysis showed that no linguistic regularity can be observed in terms of pragmatemes. However, it pointed to some linguistic characteristics being more common than others across all studied languages, such as compositionality, non-imperativeness, non-elliptical form, non-question form, and the length of fewer than five words.

The third research question concerns pragmatemes in translation, specifically within one type of audiovisual translation, namely subtitling. Existing literature lacks research that specifically examines instances of the translation of pragmatemes in various texts, with most studies focused on translating pragmatemes from one language to another individually (e.g., Sułkowska 2023), as if for a bilingual dictionary. This study aims to fill this gap by presenting an analysis of pragmatemes and their translations in context. Given the time and space constraints of this dissertation, this study focuses on one type of pragmatemes: charged units. Charged pragmatemes were chosen for this research due to their nature, which is ambiguous without any context, potentially posing particular challenges for translators. Furthermore, subtitling was selected as a type of translation that imposes specific constraints (such as the number of characters on the screen and the time for displaying a subtitle) and that takes into account information from various semiotic codes, including the visual image (Gambier 2014).

English charged pragmatemes and their translations into French and Polish were analyzed with the use of three typologies of translation techniques, each representing a different perspective (Molina and Hurtado Albir 2002, Hejwowski 2015, Díaz-Cintas and Remael 2007). The analysis revealed several methodological problems associated with the application of these typologies; however, the main observation concerned the fact that in none of the applied approaches, context was taken into account. Therefore, a new set of translation techniques in subtitling is proposed. This approach is based on context and comprises ten main techniques: deletion (of four types: context-based, technical constraints-based, repetition-based,

and visual-based), ostensibly redundant rendition, compression (of two types: syntactic and contextual), erroneous equivalent, consistency equivalent, contextual interpretation, idiosyncratic addition, creative rendition, ostensible idiomatic equivalence, equivalent. The analysis of a few chosen examples was supported with screenshots, recognizing that in audiovisual translation, visual and linguistic codes influence each other and can mutually help in understanding linguistic and visual elements (Tomaszkiewicz 2006: 80). With many interesting observations regarding every proposed technique, the equivalent technique proved to be the most prevalent in the translation of charged pragmatemes. The equivalents found in French and Polish translations of English charged pragmatemes were then subjected to the same linguistic analysis as the previously mentioned English, French, and Polish source language pragmatemes. The findings supported some of the common traits of pragmatemes, with equivalents also being mostly non-questions and non-imperative, among other characteristics.

Both the contrastive and the translational analyses provide a fresh perspective on pragmatemes, but the findings are not limited to this topic. The method of collecting corpora from Netflix captions may serve as an inspiration for other researchers, as this streaming platform offers a substantial resource of language material, providing corpora of spoken language, which is a type of corpora that has always been challenging to collect (Mikhailov, Tommola, and Isolahti 2010). Furthermore, the proposed context-based typology of translation techniques in subtitling may find application beyond pragmatemes, as its primary focus in on the context surrounding a given translation rather than a specific type of the studied unit. In terms of the main topic, i.e., pragmatemes, this dissertation suggests several interesting avenues for future research, such as studying monolexical pragmatemes, which in this study, were found to be more common than it had been previously claimed in literature (e.g., Blanco and Mejri 2018), among other potential areas of investigation.

This thesis is structured into four chapters. Following the general introduction, Chapter 1 discusses the theoretical background necessary to understand the topic. It covers definitions of key terms such as pragmateme, formulaic language, audiovisual translation, and subtitling, among others. Furthermore, the literature review presented

in Chapter 1 highlights the research gap in terms of a contrastive view on pragmatemes and their examination in translation, particularly in audiovisual translation.

Chapter 2 discusses the methods used in this study. Firstly, the content of the corpora is presented, along with the description of the tool used to collect the corpora (i.e., Language Reactor). Secondly, the method behind creating the initial list of pragmatemes from various sources is explained. Thirdly, corpus linguistic software, Unitex, and its role in searching for pragmatemes in the corpora is discussed. Finally, Chapter 2 details how the found pragmatemes were assembled and organized in the form of a table for further analysis.

Chapter 3 covers linguistic analysis of the pragmatemes found in English, French, and Polish corpora. A separate subchapter is devoted to each analyzed linguistic trait. Furthermore, Chapter 3 also presents the analyses of the discussed linguistic traits performed with the use of statistical means such as chi-squared test, probability ratio, and multiple correspondence analysis.

Chapter 4 delves into the translation analysis of charged pragmatemes. Firstly, the analysis with the use of the most general approach among the translation techniques typologies applied in this study, i.e., the dynamic and functionalist approach suggested by Molina and Hurtado Albir (2002), is presented. Then, the application of Hejwowski's (2015) model of translating idioms is discussed. The analysis of the units with the typologies already suggested in the literature ends with the application of Díaz-Cintas and Remael's (2007) model of translating cultural elements in subtitling. Finally, the issues regarding the used typologies are addressed, and a new set of techniques, representing the context-based approach, is proposed and used to analyze the studied translations.

Following Chapter 4, conclusions of the analyses are presented, along with a discussion on the study's limitations and prospects. This thesis also includes an exhaustive list of references and four appendices, including the index of pragmatemes found in the English, French, and Polish corpora, a full list of probability ratios of the analyzed linguistic characteristics of pragmatemes, a list of figures, and a list of tables.