CROSSROADS

A Journal of English Studies

ISSUE 45
2/2024

An electronic journal published
by The University of Bialystok

CrossRoads

A Journal of English Studies

ISSUE 45

2/2024

An electronic journal published by The University of Białystok

CROSSROADS
A Journal of English Studies

Publisher:
The University of Białystok
The Faculty of Philology
ul. Liniarskiego 3
15-420 Białystok, Poland
tel. 0048 85 7457450
crossroads@uwb.edu.pl
https://czasopisma.filologia.uwb.edu.pl/index.php/c/index

e-ISSN 2300-6250
The electronic version of Crossroads. A Journal of English Studies is its primary (referential) version.

Contents

Articles

Work in progress

129 YURII CHYBRAS

Phonetics and phonology of sound perception in a changing system

AIDA ALLA[1]
AAB College, Kosovo
https://orcid.org/0000-0002-3014-0667

DOI: 10.15290/CR.2024.45.2.01

An invisible storyteller or a loud recreator? A translator-centered approach to the translation of children's literature

Abstract. This paper aims to demonstrate that, like the original author, a translator of children's literature (hereafter CH. L.) possesses a distinct style or idiolect, shaped by both linguistic and extralinguistic expectations. The study focuses on the first three books of the Harry Potter series by J.K. Rowling, translated into Albanian by Amik Kasoruho, who is renowned for his contributions to the translation of classic adult literature. Given the study's scope, the analysis will concentrate exclusively on Kasoruho's creative use of the lexicon in the Albanian translation. Both internal and external factors are considered to identify and analyze translator Kasoruho's idiolect at the lexical level. Internally, sentences containing words and phrases with common patterns (e.g., archaic terms, dialectal expressions, phraseological units, substandard words) are selected from the target text. These are compared with their counterparts in the source text to determine whether such patterns reflect the author's style or the translator's linguistic preferences. Externally, these lexical clusters are assessed against the norms of children's literature translation (Ch. L. T.) to ascertain whether the translator adhered to or deviated from these norms. The findings suggest that the translator's linguistic idiosyncrasies significantly influence the translation process.
Keywords: Amik Kasoruho, children's literature translation, idiolect, lexical level, norms, translator-centered approach.

1. Introduction

Thus far, considerable research has been carried out into children's literature translation through a variety of topics and methodological angles. Klingberg (1978) paved the way for studies in Ch. L.T. by founding the scientific journal *International Research in Children's Literature*. Ever since researchers have raised numerous concerns and introduced concepts that have significantly enhanced the status and scientific credibility of this discipline within the academic community.

1 Address for correspondence: Department of English Language, Faculty of Foreign Languages, AAB College, Elez Berisha St, No. 56, 10000, Prishtina, Kosovo. E-mail: aida.alla@aab-edu.net

To mention but a few, Shavit states that "the behavior of translation of children's literature is largely determined by the position of children's literature within the literary polysystem" (1986, pp. 111–112). In other words, the adaptions and manipulations of the source text occur due to the peripheral position that this literature occupies in the polysystem.

Furthermore, Oittinen introduced the translator-centered approach to the study of the translation of children's literature. In her book *Translating for Children* (2000), she emphasizes the choices of translators and criticizes the fact that they are considered invisible. She questions the traditional approaches to translation that emphasize abstract structures of equivalence. Instead, she believes that the translator's choices are as significant as those of the original author and are often influenced by external factors such as the intentions of publishing houses and the audience. In this respect, Lathey (2016) expands on a spectrum of topics including narrative communication with the child reader; translation of the visual, dialogue, dialect, and wordplays, and retranslation to factors influencing children's publishing and globalization. Her topic proposals have been the starting point of a lot of contemporary studies to be conducted by other researchers.

This paper intends to contribute to the domain of research in Ch. L. T. by examining the linguistic peculiarities of the translator in contrast to the source text and other external factors such as the norms of Ch. L. T. This investigation will focus the analysis on a translator who has translated adults' literary books throughout his life and he embarks on translating children's literature for the first time. More specifically, the study attempts to answer the question: Is the translator an "invisible storyteller" (Lathey, 2010, p. 5) or a loud recreator? The answer to this question will be determined by further investigating the following sub-questions:

1. Do the clustered lexical features found in the corpus reflect the style of the source text author, or do they stem from the linguistic preferences of the translator?
2. Do the translations of the clusters adhere to the norms of children's literature translation (Ch. L. T.)?

2. The norms in Children's Literature Translation versus the translator's linguistic individualization (idiolect)

For this study, the norms of Ch. L. and ideologies will be revisited to analyze whether translator Kasoruho considered them as he translated the Harry Potter saga. The translation process is governed by some linguistic and extra-linguistic norms and ideologies that translators should presumably consider as they embark on the translation task. As Stephens puts it "A narrative without an ideology is unthinkable" (1995, p. 853). According to Lefevere, two factors determine the image of a literary work designed in a translation: "The translator's ideology (whether he/she willingly embraces it or whether it is imposed on him/her as a constraint by some form of patronage), and the poetics dominant in the receiving literature at the time the translation is made" (1992, p. 41). "In practice, the recreators of texts need to be aware of the ideologies, norms, conventions, and poetics, prevailing in each literature, society, and culture" (Dybiec-Gajer et. al., 2020, pp. 13–14).

For instance, when the source text is characterized by the presence of many substandard terms and unbridled style or idiomatic expressions, should the translator preserve these formats in the language of translation or adapt them following the ideologies and educational norms of the target culture? According to Desmidt (2014, p. 84), norms can be "didactic, pedagogical and technical" aside from some other more general norms such as "source-text related norms, literary aesthetic norms, and business norms" (Desmidt, 2014, p. 84), ideological, religious, et cetera, and can determine what can be translated, how, where, and when. Additionally, they can change over time, across languages, through different cultures, and from one generation to the next. Norms are used as an external benchmark because the semantic and stylistic choices made by the translator are often less pedagogical than those required by the norms of children's literature.

Regarding children's literature, the role of norms and ideologies is complicated due to several factors:

1. The need to adapt language, make abridgments, and purify content for the child reader (Klingberg, 1986);
2. The dual nature of children's literature as both a literary and pedagogical system, which leads to standardization and reflects ideological and social norms (Hunt et al., 2006);
3. Its peripheral status within the literary system, which allows for adaptations and free translations (Shavit, 1986);
4. The asymmetrical communication chain throughout the translation process (Reiss, 1982); among other factors.

O'Sullivan discusses the influence of norms in Ch. L.T., among other topics. She asserts that children's literature is "a literature into which the dominant social, cultural, and educational norms are inscribed" (2004, p. 193). Consequently, even in the most liberal countries, children's literature often reflects perspectives on language, culture, and ideology in addition to its aesthetic elements. Similarly, Van Coillie questions whether these limitations affect the exposure of the source text's cultural content in the target text (2020). He is concerned that "translations may limit diversity more than they stimulate it" (2020, p. 143), citing external factors such as "globalization and commercialization" (Van Coillie, 2020, p. 143). However, in contrast to the imposing norms mentioned above, there is also the translator's individual creativity, language adaptation, and personal style. Lathey states that:

> A degree of stylistic and semantic creativity is essential to the successful translation of texts for adults or children. Invisible storytellers have enriched the English language through an intense engagement with a source language, creating memorable phrases that have become part of children's literature. (2010, p. 6)

Translators writing for a child readership adopt translation strategies to conform to or challenge contemporary constructions of childhood. In this context, Coldiron (2012, p. 196) states

that "it is high time we set the translators free of such constraints and let them reveal the work to the reader as part of the aesthetic pleasure of the text." Re-creators of an original story, such as translators and illustrators, have specific purposes that influence how they interpret and present various elements of the text. Lathey states:

> Childhood is, after all, a volatile concept, changing its boundaries and social position according to adult requirements. Translators, too, have opinions on education and childhood reading: some have demonstrated greater autonomy than the stereotype of the hack translator allowed by their status as educators, leading literary figures, or as children's authors who were also translators (2010, p. 6).

Van Coillie supports the translator's right to make free choices, noting that "Every translator has to make choices about staying close to the source text and adapting it for a new audience" (2022, p. 144). In this respect, the translators of Ch. L. can be independent enough and be recreators of the source text in that they can set themselves free and give the text aesthetic nuances and preserve the meaning at the same time, as well as defy the norms of children's translation, which to some extent can be consequential to the final product.

3. On the studies of the translator's style

Reviewing previous studies on literary translation, it is clear that many have focused on distinguishing the translator's style from that of the source text. Mona Baker pioneered the exploration of how the "voice" of the translator can be analyzed. Baker proposed that such studies "must focus on the manner of expression that is typical of a translator, rather than simply instances of open intervention" (2000, p. 245). In other words, it is the study of the translator's "preferred or recurring patterns of linguistic behavior" (Baker, 2000, p. 245) which do not necessarily correspond to the author of the source text. An extensive collection of papers on the translator's style can also be found in the book *Style in translation: A corpus-based perspective* by Libo Huang (2015). The author presents different methods and approaches to investigating translators' styles by incorporating stylistic, rhetorical, narrative, and linguistic views. Similarly, Tim Parks in his book *Translating style: A literary approach to translation – a translation approach to literature* (2014) introduces a series of papers on translation style from the perspective of literary studies in the works of James Joyce, Samuel Beckett, Barbara Pym, etc., with a focus on the preservation of nuances of meanings in the target texts. Similarly, uniqueness in style has been vastly researched by Khatib and Al-qaoud (2021), who conducted a comparative study on authorship verification of pragmatic texts, and Koppel and Winter (2013), who investigated the identification of similarities of different materials translated by the same author.

Concerning studies of translator styles in Ch. L, Čermáková (2018) investigated the translator's choices by relying on corpus-linguistic techniques, namely keywords and cluster repetitions

in the *Harry Potter* and the *Winnie the Pooh* books. This study concluded that in most cases, repetition was replaced with synonyms due to the stylistic norms and semantic conventions of the target language, namely the Czech language. Anette Øster (2014) conducted a study on the key characteristics of the translation of Hans Christian Andersen's fairy tales. Her study revealed that the target text was closer to the folk elements than the source text. According to her, the translator's style was determined by the conventions of fairy tales, the translator's child's image, and the translator's awareness of the genre of fairy tales.

4. Challenges and competencies of the translator of Ch. L.

Although the translation of literary texts is generally considered "sensitive" due to the special use of language, as a result of the author's motivated choices, the translation of children's literature is characterized by some features, which make the work of the translators challenging and put them in dilemmas during the decision-making process. One obstacle that might condition the choices of the translator is the lack of linguistic and extralinguistic knowledge of the child reader, as stated by several authors. For more than two decades, the perception that child readers lack linguistic and extralinguistic knowledge has been challenged and reversed. Children have transcended cultural boundaries due to technological advances, frequent travel, globalization, increased cultural exposure, and translations. Two decades ago, O'Sullivan noted:

> The notion of today's children around the world, who perceive the world as a place without borders, with their books that easily cross all language and political barriers, is not something new in the academic discourse of children's literature (2004, p. 13).

Regarding the requirements and dilemmas during the decision-making process, there are many questions that the translators ask themselves during the translation process. "Translating for children has its methodology as well as a set of requirements that must be met" (Qafzezi, 2014, p. 116). Dilemmas also arise from the fact that children's literature is part of both the pedagogical and literary-aesthetic systems (Hunt, 2006). Translators face the choice of whether to preserve the linguistic and extralinguistic features of the source text, thus favoring what Venuti terms "foreignization" (Venuti, 1995, p. 19), or to adapt it to the target language, thereby "domesticating" it (Venuti, 1995, p. 19). This challenge becomes even greater when during the translation process we have interventions and expectations from other actors involved in the decision-making process starting from publishers, parents, teachers, etc.

Translators of children's literature must be well-versed in the source culture, as young readers' cultural experiences may differ significantly from the author's perspective, requiring careful adaptation. Despite the misconception that translating children's literature is simpler than translating for adults, it presents unique challenges. As Lathey (2016) notes, both the writer and the translator must grasp the nuances of language from a child's perspective. The translator should therefore integrate elements of the source culture while ensuring that the text remains

accessible and engaging for young readers, avoiding excessive foreign details that could impede comprehension.

Newmark also advises that the translator should consider the reader when translating. He states that "Taking into account the diversity of language (idioms of different characters) found in the original text, the translator tries to characterize the readers of the original and then those of the translation and decide how much attention should be paid to the readers of the translated text" (Newmark, 1988, p. 13).

Another competence of the translator is creativity. Like the author of a work, the translator must be creative and achieve a similar effect on the reader. "This creativity is even more necessary when it comes to fantasy books or humorous events. In these cases, the translator does a free translation, which allows him greater opportunities to give space to creativity and word plays" (Van Coillie, 2006, p. 135). Gillian Lathey also touches upon the need for creativity which, according to her, "requires a writer or a translator to have an understanding of the freshness of language to the child's eye and ear" (2016, p. 8). Considering the above-mentioned theoretical grounds, it can be stated that the translation of children's literature puts the translator in many dilemmas.

5. Selection of the corpus

The choice of J.K. Rowling's *Harry Potter* series for this study is based on three key reasons. First, it revolutionized children's literature, "being the first since *Charlotte's Web* to appear on the New York Times bestseller list" (Black, 2003, p. 238). Its original and translated versions have been widely researched. Lathey states, "There is a rapidly decreasing interval between publication and the worldwide translations of best-selling children's fiction such as the Harry Potter series, and publishers pay keen attention to the potential sale and licensing of world rights" (Lathey, 2010, p. 202). Second, the series blends multiple genres and styles, featuring rich intertextuality, cultural references, and poetic elements (Alla, 2017). Third, it is translated by Amik Kasoruho, renowned for his creative translation of over 60 English classics into Albanian, including *Wuthering Heights*, *Atlas Shrugged*, *The Scarlet Letter*, *Rebeka*, etc.

6. Methodology

To showcase translator Kasoruho's idiolectic features, three of the "Harry Potter" books have been taken into account: *Harry Potter and the Philosopher Stone*, (1997), (hereafter HPPS), *Harry Potter and the Chamber of Secrets*, (1998), (hereafter HPCS), *and Harry Potter and the Prison of Azkaban*, (*1999*), (hereafter (HPPA) and their Albanian equivalents *Harry Potter dhe Guri Filozofal*, (2021) (hereafter HPGF), *Harri Potter dhe Dhoma e të Fshehtave*), (hereafter HPDF), (2002) and *Harri Potter dhe Burgu i Azkabanit*, (hereafter HPBA), (2003). I chose the very first three books of Harry Potter for two reasons: 1) they are aimed at a child audience considering that the target readership is the child reader compared to other forthcoming books of the saga, which address a more mature readership (shifting into crossover literature), 2) they genuinely represent the

author's style before the saga became globalized and commercialized to the extent that the style might have been compromised.

The study focuses on the translator's lexicon. The sentences with words and phrases sharing common patterns (for example, archaic words, dialect words, phraseological units, substandard words, etc.), have been selected from the target text and then compared to their counterparts in the source text in the tables below to observe whether such patterns reflect the author's style or the translator's linguistic preferences (idiolect). For example, if the translator uses an idiomatic expression in the target text, did they do so because the author used the same expression in the source text, or was it a choice made by the translator? This study will examine the translation solution to the external norms and conventions of children's literature to determine if the translator's choices align with generally accepted linguistic tendencies. It is important to note that this paper will not quantitatively analyze the frequency of such patterns. Instead, it will provide a qualitative analysis of selected examples to illustrate the translator's unique and individual style. These examples will be organized into tables for subsequent textual analysis.

7. Analysing the translator's idiolectic features

Differences between speakers cannot be determined solely by geographical divisions or belonging to a particular ethnic group or social class. Linguistic variations cannot end with dialects. Modern linguistics recognizes that no two speakers use language in the same way. "We all have our linguistic mannerisms and stylistic idiosyncrasies, and the term reserved for an individual's special unique style is idiolect" (Simpson, 2004, p. 102). The idiolect of the translator of children's literature has rarely been studied in research. Through this paper, I aim to emphasize the fact that a translator-centered approach might bring forth extremely interesting facts such as:

1. How intricate children's literature can be,
2. How the translator of children's literature can demonstrate idiolectic features regardless of internal and external factors.

7.1. The use of words with emotionally charged suffixes

One of the stylistic features of translator Amik Kasoruho is the use of lexical chunks with connotative nuances, which carry subjective viewpoints of the author about the characters of the book. Such features make the discourse more vivid and more similar to a real-life situation.

A lexical method for conveying the connotative layers of a term involves the use of adjectives with suffixes that impart positive or negative meanings. These suffixes add emotional nuance to the word and reflect the author's attitude toward the referent or object being described. Such contrasts can express various connotations or emotional nuances, including flattering, approving, admiring, as well as pejorative, negative, offensive, and aggravating tones. These nuances contribute not only to the stylistic values of the text but also to its morpho-semantic dimensions.

Table 1. Examples of words with emotionally colored suffixes as an idiolectic feature of the translator

Examples from the source text	**Examples from the target text**
Flight of the *Fat Lady* (HPPA: 148)	Arratisja e zonjës *trashaluqe* (HPBA: 109)
He was a *big, beefy man* with hardly any neck. Mrs. Dursley was *thin* and blonde and had nearly twice the usual amount of neck. (HPPS: 1)	Ishte *madhosh, dërdëng*. Zonja Dërsli ishte *thatime*, flokëverdhë dhe me një qafë të gjatë pothuaj sa dyfishi i një qafe zakonshme. (HPGF: 7)
Wandering around at midnight, *Ickle Firsties*? Tut, tut, tut. Naughty, naughty, you'll get caught. (HPGF: 170)	Po i vini rrotull kështjellës në mesnatë o *bylmeza*? Ha, ha, ha! O *teveqela*, kanë për t'ju përjashtuar! (HPGF: 131)
Ickle Firsties! What fun! (HPPS: 138)	Na qenkan *buzëqumështit e vitit të parë.* Sa bukur! (HPGF: 106)
What looked like a *fat* little monk (HPPS: 122)	Ai që i përngjiste një murgu të shkurtër dhe *buçkan.* (HPGF: 96)
Yes, yes, you're right, of course. But how is the *boy* getting here, Dumbledore? (HPPS: 14)	Po … po, ju keni të drejtë, natyrisht. Po në ç›mënyrë do të vijë këtu *djalka*? (HPGF: 14)
…except it and watch Percy chase Fred and George all over Gryffindor tower because *they'd* stolen his prefect badge. (HPCS, 219)	…po te mos ishte se donin të shihnin Persin që po ndiqte me vrap nëpër kullën e Grifartit Fredin dhe Xhorxhin, sepse dy *horucët* i kishin rrëmbyer distingtivin e tij si prefekt. (HPGF, fq. 167)
Hagrid was *only a boy*, but he cared for me. (HPCS: 293)	Hagridi asi kohe ishtë një çunak, por u kujdes për mua. (HPDF: 224)
He caught *that thing* in his hand after a fifty-foot dive. (HPPS: 162)	E rroku *topthin* vetëm me një dorë… (HPGF: 124)
They swilled the *dregs* around AS Professor Trelawney had instructed, then drained the cops and swapped them. (HPPA: 110)	I rrotulluan *fundçet* e filxhanit, si u kishte thënë profesoresha Trelaunej, pastaj i përmbysën dhe i shkëmbyen. (HPPA: 82)
"Not lost are you, *my dear*?" said a voice in his ear, making him jump. (HPCS: 56)	Mos ke humbur gjë rrugën, *djalosh*? – I pëshpëriti në vesh një zë dhe ai kapërceu përpjetë. (HPDF: 48)
Griphook whistled and *a small cart* came hurtling up the tracks toward them. (PHPS: 80)	Unçi-unçi fishkëlleu dhe një *vagonetë e vogël* erdhi drejt tyre duke nxjerrë një rropamë hekurash që fërkoheshin. (HPGF: 64)
with their *little* plastic toys (HPPS: 218)	përmbanin ca lojëra *të vockla* plastike (HPGF: 166)
Little *tyke* wants his money's worth, just like his father. (PPPS: 23)	Ky *maskaruc* i vogël i do të tëra ç'i takojnë., deri më një, bash si i ati. (HPGF: 22)
He caught *that thing* in his hand after a fifty-foot dive. (HPPS: 162)	E rroku *topthin* vetëm me një dorë… (HPGF: 124)

Examples from the source text	Examples from the target text
They swilled the *dregs* around AS Professor Trelawney had instructed, then drained the cops and swapped them. (HPPA: 110)	I rrotulluan *fundçet* e filxhanit, si u kishte thënë profesoresha Trelaunej, pastaj i përmbysën dhe i shkëmbyen. (HPPA: 82)
"Not lost are you, *my dear*?" said a voice in his ear, making him jump. (HPCS: 56)	Mos ke humbur gjë rrugën, *djalosh*? – I pëshpëriti në vesh një zë dhe ai kapërceu përpjetë. (HPDF: 48)

The examples in italics in Table 1 are stylistic markers that describe the characters' exterior appearance. The adjective *trashaluqe*, which is formed by the root (*trasha +the suffixes luq +e*[2]) is an offensive word for *fat* in Albanian and serves as a stylistic tool conveying the point of view of the author about the character in contrast to the denotative term *e bëshme,* which means *chubby* in English. In the other example, there is the opposition of the adjective *dërd-ëng*, (*fat*) with the adjective *that-ime (thin)*. There are two antonyms, which coincide in register and style. Again, we have two lexemes with negative connotations, through which the author gives additional information to the reader about the characters in question. Mr. and Mrs. Dursley, who appear later in the story, are wicked and ruthless characters and the way they are described at the beginning of the novel foreshadows their characters.

The translator has provided two different equivalents for the term *Ickle Firsties*, which is used by Peeves to address first graders mockingly. In the first translation, the word *bylmeza* is used, which is a neologism of the translator composed of the stem *bylmet* (dairy milk) which semantically evokes the feeling of mocking, which means: *he still has milky lips*, so they are still small. For the second time, the term *Ickle Firsties* is translated as *buzëqumështit e vitit të parë* (the first year's milky lips). The mocking nuances of the phrase have been conveyed in the target language in both forms, showing the emotional state of the speaker. From the translation viewpoint, the repetition of the same term is avoided, making the translated variant more colorful.

The following example illustrates a term that fits the context and the participants in the discourse. The personal pronoun *they*, referring to the boys Fred and George, is replaced with *horucët*, derived from *horra* (*blackguard*) with the suffix *-uc* (*hor-uc-ët*). The suffix *-uc* generally adds a flattering nuance, though it can also convey negative connotations depending on the context. Similarly, the term *tyke* and its translation demonstrate a stylistic charge. Uncle Vernon, commenting on his spoiled son's unpleasant behavior, uses *tyke*, which, according to the *Oxford Advanced Learners' Dictionary* (2010, p. 1834), refers to a small child, particularly a cheeky or mischievous one, or a vulgar and uneducated person. In the target language, this is translated as

2 Albanian, as a synthetic language, is highly inflectional. There are two sufiexes added to the root of the word: the derivational morpheme *luq*, which vests the connotative/pejorative meaning to the root of the word, we also have the inflectional morpheme *e* which renders the noun into the feminine gender.

maskaruc, combining *maskara* (*blackguard*) with the diminutive suffix *-uc* (*maskar-uc*). This suffix imparts a nurturing tone to the translation, aligning with the context. Further on, in Table 1, we notice the use of nouns and adjectives with diminutive suffixes: çun-ak *(çun,)* (back-translated as *little boy*) *djal-osh* (*djalë*) back-translated as 'little boy', *top-thin* (*topi*), back-translated as 'a small ball', *fund*-çet, *(fundet) back-translated as* 'the bottoms', *të voçkla*, (*të vogla*) back-translated as *tiny, vagon-etë e vogël* (*vagon*) back-translated as 'small wagon'. These examples prove how attentive the translator Kasoruho is in conveying the semantic and stylistic layers, making the reading even more enjoyable.

7.2. The use of synonymy with emotional overtones

Synonymy is another lexical approach for conveying connotative subtleties. Generally, synonyms are words with similar meanings but subtle differences in nuance. For example, the adjectives *good* and *wonderful* express different degrees of qualification; using wonderful instead of good reflects the author's perspective on the referent (Thomai, 2005, p. 152). Synonyms also differ in their distribution and frequency of use across various fields, speaker circles, social strata, discourses, and styles. One word may be used more frequently in certain contexts, while its synonym is used less. This category includes dialectal words, which, unlike standard language, introduce nuances of colloquial speech and reveal the geographical location and social status of a character.

Additionally, synonymy may be employed to find an approximate term in the target language when an exact equivalent does not exist or when the text or word lacks specific significance, particularly with adjectives and quality adverbs. Beyond these cases, synonymy can also serve stylistic purposes, such as avoiding linguistic monotony or creating contrasts between two referents or realities.

Table 2. Examples of synonyms as an idiolectic feature of the translator

Examples from the source text	**Examples from the target text**
These fantastic party *favors* were nothing like the feeble Muggle ones the Dursleys usually bought. (HPPS: 218)	Ato *këllcaca të hatashmë* nuk kishin asgjë të përbashkët me ato *fishekzjarrët pa pikë vlere* të babanacëve. (HPGF: 166)
Piers, Dennis, Malcolm, and Gordon were all big and *stupid*, but as Dudley was the biggest and *stupidest of the lot*, he was the leader. (HPCS: 33)	Persi, Denisi, Malkolmi dhe Gordoni ishin trupmëdhej, të shëndoshë dhe *teveqelë*, por meqë Dadli ishte më i bëshëm dhe *më rrotë se të tjerët*, kryetar ishte ai. (HPGF: 30)
They had been murdered, murdered by the most feared Dark *wizard* for a hundred years, Lord Voldemort. (HPPA: 7)	Qenë vrarë nga *shtrigani më i tmerrshëm* i njëqind vjetëve të fundit, nga Fluronvdekja. (HPBA: 10)

According to the *Oxford Advanced Learner's Dictionary* (2010, p. 443), the word *favor* can refer to decorative items, such as colored paper sticks or hats, distributed to guests at a party. This sentence contrasts two realities that Harry experiences very differently: gifts in the magical world versus those in the human world. Translator Kasoruho linguistically illustrates these two realities by using distinct terms for the same concept within a sentence. For *fantastic favors*, he uses *këlcaca të hatashme* (amazing fireworks), while *the feeble Muggle ones* is translated as *fishek-zjarre pa pikë vlere* (worthless firecrackers). By employing synonyms, or "intuitive equivalents" (Čermáková, 2018, p. 126), the translator creates a dichotomy that accentuates the distinction between Harry's two worlds.

In the second example, the adjective *stupid* is translated differently in two instances. The first use is rendered as *teveqelë* (doltish) and the second as *më rrotë se të tjerët* (more knuckleheaded than the others). The translator's choice to use colloquial slang in both instances aligns with Kasoruho's idiolect, which often incorporates colloquial terms to convey a character's emotional state. This choice vividly emphasizes Harry's strained relationship with Dudley and his friends, marked by mutual hatred and disgust. The third example presents a synonym for *wizard*, which is usually translated as *magjistar* (*magician*) in other contexts. However, in this instance, *wizard* is translated as *shtrigan*, specifically referring to the malevolent sorcerer Voldemort, who killed Harry's parents. The term *shtrigan* carries a negative connotation in Albanian that effectively contrasts the benevolent magicians with the evil Voldemort.

7.3. The use of phraseological units

Another distinctive feature of translator Kasoruho is his extensive use of phraseological expressions, particularly in dialogic discourse and other contexts, which livens up the text. According to Jani Thomai, "Words and phraseological units with discourse and stylistic value, which carry different emotional nuances, possess great expressive power and impart a figurative and highly expressive quality to communication" (Thomai, 2005, p. 293). This may have been the translator's intention in choosing these phraseological units.

Table 3. Examples of phraseological units as an idiolectic feature of the translator

Examples from the source text	Examples from the target text
"*Shove off,* Malfoy," said Ron, whose jaw was clenched. Did the scary old Dementor *frighten* you too, Weasley? (HPPA: 92)	*Hiqu qafë,* Mallfoi – tha me zë të shtrënguar Roni. Edhe ty, Uesli *ta kalli datën* ai Marrosësi i shëmtuar. (HPBA: 70)
"*She deserved it*," Harry said, breathing very fast. "*She deserved what she got.* You keep away from me. *I've had enough."* (HPPA: 32)	*E bëri hak,* i tha duke marrë frymë më vështirësi. *E bëri me të vërtetë hak.* Mos m'u afro! Po iki, – tha. *Më ka ardhur në majë të hundës.* (HPBA: 28)
It was even worth being with Dudley and Piers to be spending the day. (HPPS: 26)	Dhe *e vlente barra qiranë* ta kalonte një ditë me Dadlin dhe Persin. (HPGF: 25)

Examples from the source text	Examples from the target text
The rest of them were all quite happy to join in Dudley's favorite sport: *Harry Hunting.* (HPPS: 33)	Gjithë të tjerët bashkoheshin gjithë qejf me 'te për të ushtruar së bashku sportin e tij më të dashur: *t'i binin në qafë Harrit*. (HPGF, 30)
"I look to the Prefects, and our new Head Boy and Girl, *to make sure* that no student runs foul of the Dementors," he said. (HPPA: 97)	Kam besim se Prefektët dhe Kryeshkollaret e Kryeshkollarët *do të bëjnë* çmos që askush të mos u kundërvihet Marrosësve, – tha ai. (HPBA: 73)
Bet you five galleons the next one *dies*. (HPCS: 282)	Vë bast se ai që do ta ketë rradhën ketej e tutje sot *do të kthejë këmbët nga dielli*. (HPDF: 215)

The words in italics in the examples in Table 3 are neutral terms taken from the English lexicon. However, their counterparts in the target language are phraseological expressions with more intense connotations: *happy* – gjithë qejf (full of joy); *Harry Hunting – t'i binin në qafë Harrit* (Let's take on Harry); *horrible – për dreq* (like hell). The verb *dies* in the source text is translated as *do të kthejë këmbët nga dielli* (kicked the bucket), which has a humorous effect on the child reader and makes the communication more casual. This humorous effect is consistent across all the examples in the table.

7.4. The use of archaic words

The language system extends beyond its representation at any given time (Lloshi, 2005, p. 36). The translator's deliberate use of obsolete words, particularly those from everyday colloquial speech, creates a distinct stylistic effect and introduces variety and additional nuances to the text. In the case of this corpus, intended for children and adolescents, obsolete words serve a pedagogical purpose: they introduce new generations to these terms, preventing their obsolescence. Many of these obsolete words are borrowed from Turkish and occupy a place in the Albanian lexicon as part of casual, colloquial discourse. Kasoruho, not a purist of the Albanian language, believes in enriching the language with borrowings from other languages. In an interview included in his book *No Grudges*, he remarked that "the persisting struggle to always resort to the Albanian lexicon is futile" (Kasoruho, 2013, p. 282). Below, I will provide examples of archaic words that highlight this idiolectic feature of the translator.

Table 4. List of examples of archaic and oriental words as an idiolectic feature of the translator

Examples from the source text	Examples from the target text
New students at Hogwarts were sorted into Houses by trying on the Sorting Hat (PHPA: 96)	Ata *që vinin sefte* në Hoguorts vinin në kokë kapëlën Folëse. (PPBA: 72)

Examples from the source text	Examples from the target text
Harry was glad school was over, but *there was no escaping* Dudley's gang. (HPPS: 33)	Harri ishte shumë i kënaqur që shkolla kishte mbaruar, *por s'kish derman* t'i përvidhej bandës së Dadlit. (HPGF: 30)
Darling, *you haven't counted* Auntie Marge's present, see, it's here under this big one from Mommy and Daddy. (HPPS: 22)	Shpirt, *s'e ke futur në hesap* dhuratën e teze Marxhit. Shikoje, ja ku e ke, nën këte dhuratë të madhe të babait dhe të mamasë. (HPGF: 22)
but Harry and Ron pretended *to be enjoying* them. (HPPS: 150)	…por Harri dhe Roni u shtirën sikur u pëlqyen *kiamet*. (HPGF: 115)
a *large* doughnut in a bag (HPPS: 4)	një *alamet* kulaçi të mbështjellë me letër (HPGF: 9)
Ron was leaning out of the back window of an old turquoise car, which was parked *in midair.* (HPCS: 25)	Roni kishte dalë jashtë nga xhami i mbrapmë i një makine të vjetër të kaltër, që ishte parkuar *në hava.* (HPDF: 25)
The falcon…my dear, you have a deadly *enemy.* (HPPA) p 112.	Këtu ka një skifter… i dashur, ti ke një *hasëm* për vdekje. (HPBA: 88).
"I realized *that*," said Harry, ducking as Hagrid *made to brush* him off again. "I told you, I was lost."(HPCS: 56)	"E kuptoj" tha Harri, duke dashur t'i shpëtonte Hagridit që po *gatitej* ta shkundëte përsëri fort. (HPDF: 49)
Aunt Petunia *had decided* it must have shrunk in the wash. (HPPS: 26)	Emta Petunia i *kishte dhënë karar* se do të kishte hyrë duke u larë në rrobalarëse… (HPGF: 25)
Dudley's mouth fell open in horror, but Harry's *heart gave a leap*. (HPPS: 23)	Dadli hapi gojën i tmerruar, por Harrit *i brofi zemra nga gëzimi*. (HPPS: 23)
"Where did you come out?" Ron asked. "Knockturn Alley," said Hagrid grimly." "*Brilliant!*" said Fred and George together. (HPCS: 58)	"Ku ishe?" pyeti Roni. "Në Nokturn Alli" tha Hagridi i ngrysur. "*Hata fare!*" thirrën njëzëri Fredi me Xhorxhin. (HPDF: 50)
The *trouble* was, there was already someone sitting on it. (HPCS: 11)	Veç *gjynah* që shtrati i tij ishte i zënë. (HPDS: 14)
Professor Trelauney was staring into the tea-cup, *rotating it anti-clockwise*. (HPPA: 111)	Profesoresha Trelaunej shikoi filxhanin, duke e rrotulluar në drejtom të kundërt me akrepat e *sahatit*. (HPBA: 83)
"*So we've heard*", said Lupin more coldly. (HPPA: 389)	E kemi dëgjuar këtë *mesele*, tha Lupini edhe më ftohtas. (HPBA: 274)
"It certainly seems so," said Dumbledore. "*We have much to be thankful for*" (HPFS: 11)	"Ashtu duket", iu gjegj ai. "*Duhet te themi shyqyr disa herë.*" (HPGF: 14)

Table 4 provides examples of archaic words of Oriental (Turkish) origin. However, this feature is absent in the source language. The omission of such terms in the original text indicates that their inclusion is a stylistic choice made by the translator, reflecting an idiolectic characteristic. It could also be argued that using archaic words might not be ideal for young readers, as they may struggle to understand them. Frequent interruptions in reading due to unfamiliar terms could lead to a loss of interest and potentially cause children to stop reading the book.

7.5. The use of dialect lexicon

Another notable idiolectic feature is Kasoruho's use of dialectal words for stylistic purposes. Before analyzing this feature, it is important to consider scholarly perspectives on the use of dialectal lexicon in literature. Jani Thomai observes that "many words and expressions from dialects can enrich the standard language, replacing foreign terms and adding exciting nuances and variety" (2005, pp. 282–283).

Kasoruho draws from both major Albanian dialects, Gheg and Tosk, but he tends to favor the Gheg dialect's vocabulary. Scholars agree that both dialects contribute equally to the enrichment of the standard language, each adding diverse registers and styles "The evaluation of a word or expression is based not on its dialectal or regional origin, but on its real value within the lexical system of the standard language" (Thomai, 2005, p. 283). This enriching process is supported by the shared elements between the two Albanian dialects. Mona Baker also discusses the role of dialects in translation, noting that they contribute to what she refers to as "the evoked meaning" (1992, p. 15).

Table 5. List of examples of dialect lexicon as an idiolectic feature of the translator

Examples from the source text	**Examples from the target text**
…tureens of *buttered* peas, silver boats of thick, rich gravy and cranberry sauce. (HPPS: 218)	…supiera me bizele të gatuara me *tëlyen*, mbajtëse salce me salca të trasha (HPGF: 165)
MRS Dursley *sipped* her tea through pursed lips. (HPPS: 7)	Zonja Dërsli *gjerbi* çajin me buzë të shtrënguara. (HPGF: 12)
Harry put his quill between his teeth and reached underneath his *pillow* for his inkbottle and a roll of parchment. (PHPA, p2)	Harri kafshoi penën dhe rrëmoi poshtë *nënkresës* se mos gjente bojën e shkrimit dhe një rrotull pergamene. (HPPA: 7)
He'd probably find himself *locked in* the *cupboard* under the stairs for the rest of the summer. (HPPA: 2)	Mbase do ta *ndrynin* harrin në *zgëqin* poshtë shkallëve për tërë pjesën tjetër të verës. (HPBA: 8)
And he *threw* the receiver back onto the telephone as if dropping a poisonous spider. (HPPA: 4)	Dhe *e flaku* tej dorezën e telefonit, sikur të ishte një merimangë nga ato të helmuarat. (HPPA: 9)

Examples from the source text	Examples from the target text
"Stop it, Oliver, you're embarrassing us," said Fred and George Weasley together, *pretending* to blush. (HPPA: 115)	Oliver, po na bën të skuqemi, – thanë me një zë Fredi dhe Xhorxhi, duke bërë *kinse* po e ndjenin veten ngushtë. (HPBA: 111)
His eyes fell on a huddle of *these* weirdos standing quite close by. (HPPS, p 3)	Shikimi shkoi mbi një trumë *asi* tuhafësh, krejt pranë tij. (HPGF: 8)
People throughout the reptile house screamed and started running for the exits. (HPPA: 413)	*Gjindja* kishte zënë të ulërinte dhe të vraponte drejt portave për të dalë jashtë. (HPBA: 291)
But Fudge was shaking *his head* with a small smile on his face. (HPPA: 413)	Por Gjelsheqeri shkundi *kryet* duke vënë pak buzën në gaz. (HPBA: 291)
Panting from the effort of dragging his trunk. (HPPA: 33)	*Duke gulçuar* nga lodhja. (HPPA: 29)
It certainly seems so, *said* Dumbledore. (HPFS: 11)	Ashtu duket, iu *gjegj* ai. (HPGF: 14)

The examples in Table 5 illustrate that the use of dialectal words is a distinct feature of translator Kasoruho's idiolect and is not intended to replace equivalent terms from the source language's dialectal lexicon. Instead, these dialectal words serve stylistic purposes by creating a contrast between standard and non-standard language. Additionally, the Gheg dialectal terms pique the curiosity of child readers due to their unusual sound.

7.6. The intentional deviations from the standard language

In conversations between characters, the reader often discerns their level of closeness or distance (tenor) through their linguistic interactions. Below are additional examples of non-standard language that suit the contextual situations in which the characters find themselves.

Table 6. List of examples of intentional deviations from the standard norm as an idiolectic feature of the translator

Examples from the source text	Examples from the target text
A fine thing it would be if, on the very day YouKnow-Who seems *to have disappeared* at last, the Muggles found out about us all.	Pikërisht ditën kur duket se më në fund Ti-E-Di--Kushi *e ka këputur qafën*, Babanacët të marrin vesh punën tone.
Well, they're not completely stupid. They were bound to notice something. (HPFS: 11)	*Epo nuk janë edhe aq leshko*. Heret apo vonë do të vinin re ndonjë gjë. (HPGF: 14)
But that's no reason *to lose our heads.*	Por kjo s'do të thotë se *duhet të humbasin fiqirin.* (HPGF: 14)

Examples from the source text	Examples from the target text
I'll bet that was Dedalus Diggle. *He never had much sense.* (HPFS: 11)	Vë bast se do të ketë qenë Dedalus Luksi. Gjithmonë *ka qenë me një vidhë mangut.* (HPGF: 14)
People are being downright careless, *out on the streets* in broad daylight, not even dressed in Muggle clothes (HPFS: 11)	Po sillen krejt pa mend, *duke u sorollatur rrugëve* pa u veshur të paktën si Babanacët. (HPFS: 14)
On the very day YouKnow-Who seems to have disappeared at last.	Pikërisht ditën kur duket se më në fund Ti-E-Di-Kushi *e ka këputur qafën.* (HPFS: 11)
He opened the back of the camera. *Good gracious! said Madam Pomfrey.* (HPCS: 190)	Hapi kapakun e mbrapmë të makinës. *Dreq! thirri Madamë Pomfri.* (HPDS: 149)
Out of the way, you, he said, punching Harry in the ribs. Caught by surprise, Harry *fell hard* on the concrete floor. What came next happened so fast no one saw *how it happened* – one second, Piers and Dudley were leaning right up close to the glass, the next, they *had leaped back with howls of horror.* (HPCS: 190)	*Qërohu, ti!* i tha d*uke ia këputur me një grusht* brinjëve Harrit, i cili, i zënë në befasi, *ra thes* në tokë. Çka ndodhi pastaj u zhvillua aq me vërtik sa skush s'e kuptoi *si e qysh*: pak më përpara Persi dhe Dadli qenë përkulur pranë xhamit dhe një çast më vonë *kërcyen mbrapa duke thirrur të lebetitur.*

Most of the translated italicized words in Table 6 are more "unprincipled" than their counterparts in the source text. For example, *Good gracious!* has been rendered into the target text as *Dreq!*, which in the source language means *Damn!; Out of the way* is rendered into the target text as *Qërohu, ti!*, that, if back-translated, would mean *Get the hell out of here!.* These examples showcase that translator Kasoruho ignored the norms of children's literature, which call for language normalization/standardization (Desmidt, 2014), and favored creative intuition with the main intention of producing a target text that is purely aesthetical.

8. Conclusions

In discussing the peculiarities of the translator's style in the Harry Potter corpus, I tried to showcase that the translator of children's literature, namely Amik Kasoruho, was a loud recreator in that he made use of his idiolectic features as well as his recreative linguistic traits to produce a target text which departed from the general models of children's literature translation. More specifically, I observed the abundant use of emotionally colored and diminutive suffixes, the use of synonymy with emotional overtones, the use of phraseological units, the use of archaic words, the use of dialectic lexicon and the use of substandard words, which resulted to have been used for stylistic reasons. I base this conclusion on the fact that these features did not manifest themselves in the source text.

In discussing the characteristics of children's literature, I highlighted several limitations and norms, such as the challenges young readers face with dialectal or outdated words and the

pedagogical nature of the genre. I observed that translator Amik Kasoruho prioritized the aesthetics of the narrative over strict linguistic accuracy. Kasoruho seemed to overlook the role of children's literature as an educational and pedagogical tool for young readers. It appears that Kasoruho adheres strongly to his idiolect and values, prioritizing the full potential of the Albanian language in translation over adherence to conventional norms or ideologies.

References

Alla, A. (2017). *Aspekte të barasvlershmërisë në përkthimet e letërsisë bashkëkohore për fëmijë nga Amik Kasoruho* [Issues of equivalence in children's literature translations by Amik Kasoruho – Unpublished doctoral dissertation]. University of Tirana.

Al-Khatib, M.A. & Al-qaoud, J. K. (2021). Authorship verification of opinion articles in online newspapers using the idiolect of the author: A comparative study. *Information, Communication & Society, 24*, 1603–1621.

Aguado-Giménez, P. & Pérez-Paredes, P.F. (2005). Translation-strategies use: A classroom-based examination of Baker's taxonomy. *Meta*, *50*, 294–311.

Baker, M. (2000). Towards a methodology for investigating the style of a literary translator. *Target. International Journal of Translation Studies*, *12*, 241–266.

Black, Sh. (2003). The magic of Harry Potter: Symbols and heroes of fantasy. *Children's Literature in Education, 34*, 237–247.

Coldiron, A.E.B. (2012). Visibility now: Historicizing foreign presences in translation. *Translation Studies, 5*, 189–200.

Čermáková, A. (2018). Translating children's literature: Some insights from corpus stylistics. *Ilha do Desterro*, *71*, 117–133.

Desmidt, I. (2014). A prototypical approach within descriptive Translation Studies? Colliding norms. In J. V. Coillie & W. P. Verschueren (Eds.), *Children's literature in translation: Challenges and strategies* (pp. 79–96). Routledge.

Dybiec-Gajer, J., Oittinen, R. & Kodura, M. (2020). *Negotiating translation and transcreation of children's literature*. Springer.

Libo, H. (2015). *Style in translation: A corpus-based perspective*. Springer.

Kasoruho, A. (2013). *Pa Mëri: Publicistika e jetës time.* [No grunges: The journalism of my life]. Pegi.

Lathey, G. (2010). *The role of translators in children's literature: Invisible storytellers*. Routledge.

Lathey, G. (2016). *Translating children's literature*. Routledge.

Lefevere, A. (2016). *Translation, rewriting, and the manipulation of literary fame*. Routledge.

Lloshi, Xh. (2005). *Stilistika e Gjuhës Shqipe dhe Pragmatika.* [Stylistics and pragmatics of the Albanian Language]. Albas.

Klingberg, G., Ørvig, M. & Amor, S. (Eds.). (1978). *Children's books in translation: The situation and the problems*: *Proceedings of the third symposium of the International Research Society for Children's Literature,* Held at Södertälje, August 26-29, 3, Almqvist & Wiksell International for the Swedish Institute for Children's Books.

Klingberg, G. (1986). *Children's fiction in the hands of the translators*. CWK Gleerup.

Koppel, M. & Winter, Y. (2014). Determining if two documents are written by the same author. *Journal of the Association for Information Science and Technology, 65*, 178–187.

Newmark, P. (1988). *A textbook of translation*. Prentice-Hall International.

O'Sullivan, E. (2004). Internationalism, the universal child and the world of children's literature. In P. Hunt (Ed.), *International companion encyclopedia of children's literature* (pp. 13-25). Routledge.

O'Sullivan, E. (2004). Comparative children's literature. In P. Hunt (Ed.), *International companion encyclopedia of children's literature* (pp. 191-202). Routledge.

Oittinen, R. (2002). *Translating for children.* Routledge.

Øster, A. (2014). Hans Christian Andersen's fairy tales in translation. In J. V. Coillie & W. P. Verschueren (Eds.), *Children's literature in translation: Challenges and strategies* (pp. 141–155). Routledge.

Parks, T. (2014). *Translating style: A literary approach to translation – a translation approach to literature*. Routledge.

Qafzezi, E. (2014). *Aspekte teorike dhe praktike të përkthimit të letërsisë për fëmijë*. [Theoretical and practical issues of children's literature translation – Unpublished doctoral dissertation]. University of Tirana.

Shavit, Z. (1986). *Poetics of children's literature*. University of Georgia Press.

Simpson, P. (2004). *Stylistics: A resource book for students*. Routledge.

Stephens, J. (1995). Writing by children, writing for children: schema theory, narrative discourse and ideology. *Revue belge de philologie et d'histoire, 73*, 853–863.

Thomai, J. (2005). *Leksikologjia e Gjuhës Shqipe.* [Lexicology of the Albanian Language]. Toena Publishing.

Venuti, L. (2017). *The translator's invisibility: A history of translation*. Routledge.

Van C. J. (2014). Character names in translation: A functional approach. In J. V. Coillie & W. P. Verschueren (Eds), *Children's Literature in Translation* (pp. 123–140). Routledge.

Van C. J. (2020). Diversity can change the world: Children's literature, translation and images of childhood. In J. V. Coillie, & J. McMartin (Eds.), *Children's literature in translation: Texts and contexts* (pp. 141–159). Leuven University Press.

Venuti, L. (1995). *The translator's invisibility: A history of translation*. Routledge.

Primary Sources:

Rowling, J.K. (1997). *Harry Potter and the Philosopher Stone.* Bloomsbury.

Rowling J.K. (1998). *Harry Potter and the Chamber of Secrets.* Bloomsbury.

Rowling J.K. (1999). *Harry Potter and the Prison of Azkaban*. Bloomsbury.

Rowling, J.K. (1997). *Harry Potter and the Philosopher Stone;* translated by Amik Kasoruho (2001) as *Harri Potter dhe Guri Filozofal.* Shtëpia Botuese Dituria.

Rowling, J.K. (1998). *Harry Potter and the Chamber of Secrets;* translated by *Amik Kasoruho* (2002) as *Harri Potter dhe Dhoma e të Fshehtave*. Shtëpia botuese Dituria.

Rowling, J.K. (1999). *Harry Potter and the Prison of Azkaban;* translated by *Amik Kasoruho* (2003) as *Harry Potter dhe Burgu i Azkabanit*. Shtëpia botuese Dituria.

***

Assist. Prof. Dr. **Aida Alla** is a lecturer at the Faculty of Foreign Languages, AAB College, where she teaches courses such as Translation Theory, Translation Practice, and Syntax. Her academic expertise lies in comparative linguistics, with a focus on English-Albanian translation studies. She completed her doctoral studies at the University of Tirana, where her dissertation focused on comparative analysis of the linguistic structures in English and Albanian, through the lens of translation equivalence. Her research examined how syntactical, stylistic, and cultural elements are rendered in literary translations from English into Albanian. In addition to her academic work, Aida Alla has extensive professional experience in translation and interpretation with international organizations in Tirana. Furthermore, she has presented her research at various regional and international scientific conferences, where she has shared her insights into translation studies.

MATEUSZ BIAŁAS[1]
University of Warsaw, Poland
https://orcid.org/0000-0002-1209-4142

DOI: 10.15290/CR.2024.45.2.02

Power bottom, *gay versatile*, *top persistant*, and other borrowings from English in erotic biographies of gay and bisexual porn stars on French adult websites

Abstract. This paper examines lexical borrowings from English in the erotic biographies of gay and bisexual porn actors available on French websites. The study adopts an anthropological perspective, drawing on David Le Breton's concept of the *'liberated' body* and Marie-Anne Paveau's typology of pornographic discourse. Three main types of borrowings have been identified: full borrowings (e.g., *bareback*), hybrid borrowings (e.g., *hardeur*), and semantic calques (e.g., *versatile*). The analysis reveals that full borrowings are predominant in the corpus, particularly in conceptual and lexical anaphors referring to specific antecedents, namely male porn stars. Moreover, these borrowings serve various functions in constructing the representations and identities of porn actors within the discourse examined, reflecting both their physical attributes, social roles, and sexual preferences. The investigation emphasizes the significance of borrowings as one of the key mechanism for enriching the French language, especially in the context of persuasive discourse, which mirrors the evolution of societies and contemporary cultural and social phenomena. Ultimately, the study may contribute to the understanding of how linguistic borrowings shape the portrayal of marginalized groups in media and influence the broader discourse on sexuality and identity.
Keywords: biography, bisexual, body, borrowing, discourse, gay, identity, pornstar.

1. Introduction

In her latest book, *La néologie de l'adjectif en français actuel* ('The neology of an adjective in Modern French'), Alicja Kacprzak (2019, p. 25) points out that French lexicology, while offering

1 Address for correspondence: University of Warsaw, Institute of Applied Linguistics, ul. Dobra 55, 00-312 Warsaw, Poland. Email: m.bialas@uw.edu.pl

several classifications of neological mechanisms, has marginalized borrowing as one of the commonest lexical enrichment procedures. Rather, the Polish researcher feels strongly that it is only due to the typologies of neologisms put forward by Guiraud (1967) and Guilbert (1975) that borrowing has been, so to speak, restored to favor among many other instruments of lexical creativity. More recently, Sablayrolles (2000), having made modifications to the classification of neologisms proposed by Tournier (1991) for the English language, created a model of lexicogenic matrices where he distinguished both the internal matrices containing a great number of lexicogenic processes (morphological, semantic, syntactic, etc.) and, on the other hand, the external matrix which includes borrowing itself. Over and above, it is worth emphasizing the fact that a lot of scientific works dealing with new borrowings are not only testament to the state of allogenic vocabulary in different languages today, but also to the evolution of societies, ethnicities, minorities, etc., for the foreign lexemes do reflect a significant number of human attitudes, experiences, relationships, behaviors, and so on. In other words, borrowings, considered as a purely neological phenomenon, accompany diverse communities in their cultural, social, and economic lives in which they seem to denote quite particular entities.

2. General overview and aims of the research

The designation of a specific entity in the mental representation that exists in the extra linguistic reality appears extremely striking, especially as to the construction of anaphoric reference in the new media of the 21st century. This paper is the result of the continuation of our research on referential expressions in external pornographic discourse where various types of anaphors seem to have become a truly common occurrence.

In the research hitherto conducted, several types of anaphors had been examined and the outcomes of our investigations into a specific corpus – erotic biographies perceived as a particular genre of external pornographic discourse – presented at international conferences in Bulgaria, Canada, Czech Republic, France, Lithuania, Poland and Romania. Our main purpose was to check how the phenomenon of reference is conceived in multimodal texts of a persuasive nature, where images, films and photos play a fundamental referential role. In the wake of our analyses, we succeeded in putting forward key anaphoric expressions referring to a specific referent: porn actors, and, as a consequence, discursive effects typical of this linguistic (diastratic) variety of the French language with its wide range of lexical and stylistic devices.

Upon closer examination, it turned out in a further study of the aforementioned corpus that in the previously distinguished anaphors, principally in some of their types, there is an interesting profusion of borrowings from the English language. Moreover, those loanwords seem to perform interesting functions (Goudaillier, 2001) within the particular slang, where they appear quite interestingly. Thus, the aim of this paper is to distinguish the types of English borrowings that are present in the biographies of gay and bisexual X-actors on French pornographic websites today.

3. Research Perspective

For our corpus analysis, we have adopted an anthropological research perspective. More specifically, we have conducted our study with reference to the concept of a 'liberated' body by Le Breton, who emphasizes the sensory exploration through diverse practices offering an unprecedented use of the human body (2017, p. 188). According to the French anthropologist and philosopher, it is crucial to note that

> the search for well-being through the best physical use of oneself, in energetic engagement with the world, responds to the need to restore anthropological roots, made precarious by the social conditions of existence. We know the growth of psychological disorders linked to the deficiencies of narcissism, the impression of feeling nothing, the inner emptiness, the staggering of the senses and intelligence, the whiteness of existence. The sensory exploration favored by sophrology, massages, yoga, relaxation, tai chi, martial arts, etc. among other practices proposing an unprecedented use of the body, reflects this anthropological need for a new alliance with an underused corporeal reality (Le Breton, 2017, p.18)[2].

4. Primary sources and research procedure

First, it is indispensable to note that the investigation into a particular type of social discourse that we carry out in this paper is based on a typology of pornographic discourse by Paveau (2014). In one of her books, the French linguist has put forward and defined its three essential types from a pragmatic point of view: discourse on pornography – that of evaluation which may aim to display defense, criticism or even violent stigmatization of non-standard forms of representation; internal discourse – that which relates to the porn production or pornographic work, namely the discursive content of the literary work, script, comic strip, work of art, etc., and external discourse – that which emanates from the publishing, or the pornographic industry and which aims to name, classify, categorize works, actors, products, etc. with a view to attracting as many recipients as possible.

Second, the discourse analysis that we propose in this paper is based on a corpus (ca. 10,000 words), for which we have analyzed a non-standard linguistic material, namely a hundred information sheets, hereinafter referred to as *erotic biographies*, which constitute a prominent example of a genre of external pornographic discourse. More specifically, we have analyzed a diastratic variety of the French language, that is a selection of ninety-three biographies of homo- and bisexual X-actors[3] gathered on the French website: `https://www.videosxgays.com/`.

---

2 All Francophone sources have been translated into English by the author of this paper.

3 Statistically speaking, an average X-actor is 31 years old, 177,5 cm tall and he weighs 74 kg. In terms of sexuality, and more specifically, in terms of sexual roles, top actors are represented by 25 cases, bottom actors by 7 cases, versatile ones by 40 cases, and 41 cases are not determined in this regard. As far as the distribution

Third, we would like to point out that, methodologically speaking, we are going to examine solely lexical borrowings from the English language. Indeed, they seem to constitute one of the most interesting neological phenomena in modern French, as depicted by many examples from the corpus studied. With this end in view, we are going to use a definition forged by Christiane Loubier (2011, p. 14), who maintains that a lexical borrowing is either "a complete borrowing (form and meaning) or a partial borrowing (form or meaning) of a foreign lexical unit". Furthermore, in order to emphasize the difference between lexical borrowings and other two main types of borrowings: syntactic and phonetic, the Canadian linguist is inclined to believe that it does primarily concern the word "in its relationship form-meaning", and therefore, she sheds light on four subtypes of lexical borrowings in her own classification:

a) **Complete (full) borrowing** – both form and meaning of the word are borrowed, e.g. *cool*, *shopping*, *brownie*;
b) **False borrowing** (mock anglicism) – a word which comprises foreign lexical units, but which is non-existent in the lending language (English), e.g. *tennisman*, *recordman*;
c) **Hybrid borrowing** – a word which contains lexical or morphological units from both languages, e.g. an English stem and a French suffix, or a compound noun consisting of both French and English words, e.g. *dopage*, *monitorage*, *chanteur de rap*, *surfeur*, *blogueur*;
d) **Calque** (loan translation) – a word or an expression in a language that is a translation of a word or expression in another language[4], e.g.
    - morphological – a word, usually compound, which is composed of native vocabulary, yet created through translation, very often-literal translation of foreign (English) terms, e.g. *finaliser* (from: *to finalise*), *gratte-ciel* (from: *skyscraper*), *lune de miel* (from: *honeymoon*);
    - semantic – a word that is used in the same meaning as in the lending language (English); yet, this word already exists in the borrowing language (French) where it has a completely different meaning, e.g. *réaliser* (from: *to realise = se rendre compte*), *digital* (from: *digital = numérique*);
    - idiomatic – is a word-for-word translation of a foreign fixed expression, e.g. *Ce n'est pas ma tasse de thé = It's not my cup of tea.*

5. Discussion of results

In light of our corpus analysis, it must be stated that three types of lexical borrowings from English have been spotted: full borrowings, hybrid borrowings, and semantic calques.

---

of porn actors by country of citizenship is concerned, a vast majority of actors come from the United Stated of America whereas the second and third positions are occupied by French and Czech actors respectively. Moreover, there is a significant representation of actors hailing from Canada, Brazil, Spain, and the United Kingdom. Among other ethnicities, our corpus is also comprised of the representatives of such countries as Hungary, Iran, Ireland, Lebanon, Mexico, Paraguay, Peru, Romania, and Russia.

4 https://www.oxfordlearnersdictionaries.com/definition/english/calque?q=calque

5.1. Full borrowings

First, it turns out that our primary sources abound in complete borrowings, which represent the majority of all the lexical borrowings (14 cases in total). It is necessary to point out, however, that most of them occur in associative anaphors – based on the conceptualization of lexical anaphors – and conceptual anaphors – a particular type of more complex anaphoric expressions that summarize or condense the content of the antecedent into an extended syntagma or even a whole sentence. In addition, it should be noted that those anaphors are sometimes equated with adjectival anaphors and our primary sources contain a plethora of such expressions in reference to specific antecedents: X-actors exclusively. The above observations can be illustrated with many concrete examples of such anaphors, all of which contain full English borrowings.

Undoubtedly, the adjective *sexy*[5] seems to be used the most frequently in the primary sources, as it may be proven by an impressive number of examples: (1) *Ce jeune mec est intelligent et* ***sexy***, *calme et sûr de lui quand il s'agit de s'éclater.* ('This young guy is intelligent and sexy, calm and self-confident when it comes to having a blast.'); (2) *Difficile de faire plus chaud et* ***sexy*** *que ce bel étalon.* ('It's difficult to be more hot and sexy than this handsome stallion.'); (3) *Sean Ford est un jeune mec bien foutu qui est aussi* ***sexy***… ('Sean Ford is a young, well-shaped guy who is also sexy…'); (4) *À la fois juvénile et* ***sexy***, *son physique fin et athlétique fait le bonheur de ses partenaires de tournage* ('Both youthful and sexy, his fine and athletic physique is the delight of his shooting partners'); (5) *(…) ses tatouages* ***sexy*** *sont de loin ce qu'on retient le plus chez lui* ('his sexy tattoos are by far what we remember the most about him'); (6) *Il est* ***sexy****, sûr de lui, et physiquement, ses formes galbées font tourner plus d'un regard.* ('He is sexy, self-confident, and physically, his shapely forms turn more than once glance.'). What is also worth emphasizing is the fact that – despite one of the canonical rules of the French grammar, stipulating that the adjective must be in concord with the noun to which it refers – the word *sexy* remains invariable and is solely acceptable in the informal register of the French language[6].

5 In the present study, the word *gay*, borrowed from American English (see: `https://www.larousse.fr/dictionnaires/francais/gay/36352`), has not been taken into account, for its numerous occurrences are manifest in this particular genre of pornographic discourse. Nevertheless, it would appear quite interesting to note that this lexeme – used in modern French as both a noun and an adjective in order to denote a homosexual person – is somewhat challenging in terms of its correct usage: even the most renowned French dictionaries have discordant views on whether it should be variable in number or not (compare: `https://dictionnaire.lerobert.com/definition/gay`). Furthermore, it seems even more interesting to see that the adjective *gay* [today's spelling: *gai*], coming from the Gothic language where it meant *impetuous*, has existed in French since the 17th century as a synonym for *sprightly*, *joyous*, *vivacious*, *ardent*, *jovial*, *agreeable*, etc. (see: `https://www.dictionnaire-academie.fr/article/A1G0063`). Yet, according to the latest edition (9th) of *Le Dictionnaire de l'Académie française*, the word in question might have derived from Old Provençal (11th c.), where it referred to someone who is exuberant, bright or sharp.

6 `https://www.larousse.fr/dictionnaires/francais/sexy/72492`

Moreover, among many other English words to be found quite abundantly in our corpus, the nouns *star*[7] and *look* are definitely not scarce. The former one, grammatically feminine in French, is always used figuratively and, needless to say, it denotes someone famous and excellent in their performance[8], e.g. (7) *Cette **star** du porno* ('This pornstar'); (8) *une porn**star** gay* ('a gay pornstar'); (9) *la **star** internationale de l'année par Prowler en 2018* ('the international star of the year 2018 by Prowler'); (10) *Cette **star** du porno arbore un joli tatouage sur son torse qui le rend facilement reconnaissable*. ('This porn star displays a nice tattoo on his chest, which makes him easily recognizable.'). The latter, nevertheless, is claimed to carry rather pejorative connotations, and is used exclusively in the sense of a manner of behaving, dressing oneself or, in other words, a general appearance of someone or something considered characteristic of such or such fashion[9]: (11) *Son **look** geek* ('His geeky look'); (12) *Il a un **look** viril et une belle gueule*, *mais ce gars*… ('He has manly looks and a handsome gob, but this dude…'), etc. Still, we find it interesting to mention that this particular meaning of the noun *look*, which has penetrated many other languages than French, stems from another type of a discourse based on persuasion: the advertising one, striving to put a positive gloss on it: "This word is often used on women magazine covers, for it is tightly connected to the world of fashion" (Kozłowska, 2021, p. 24).

Nonetheless, lest one underestimate the paramountcy of other complete borrowings from English in the corpus studied, we find it important to point out that most of them are inextricably intertwined with sexual intercourse as part of human experiences, and thus inherent in this type of social discourse. Indeed, the full borrowings that we have managed to detect in our research material also belong to the large semantic field of coitus and refer principally to the particular ways of having sex: from the softest, e.g. (13) *Se faire baiser dans le cul **bareback***[10] ('Getting fucked in the ass bareback'); (14) *Il tourne parfois des scènes **bareback** d'une rare bestialité* ('He sometimes shoots bareback scences of rare bestility'); (15) *Ce jeune Tchèque de 26 ans a déjà un*

7 In our primary sources, the word *star* also occurs as part of a proper name; more specifically, it is used as a pseudonym by one of the Brazilian pornstars: *Andy **Star** aime se faire baiser dans le cul bareback* ('Andy Star loves getting fucked in the ass bareback'), and, onomastically speaking, constitutes an interesting example of an anthroponym. Nonetheless, we would like to point out that, with regard to discourse analysis, a term *pornonym* could be applied; indeed, it is Paveau (2014, p. 140) who introduced this noun (*pornonyme*) into the class of proper names, maintaining that "the pornographic universe is (also) a universe of words and names: the pseudonyms of X-actors, X-actresses, but also of X-directors are intrinsic part of the structure of the pornographic milieu, both as industry and culture".

8 https://dictionnaire.lerobert.com/definition/star

9 https://www.larousse.fr/dictionnaires/francais/look/47777

10 The word *bareback* comes from equestrian terminology and refers to riding 'on a horse without a saddle' (see: https://www.oxfordlearnersdictionaries.com/definition/english/bareback?q=bareback), 'on the bare back of a horse' (see: https://www.merriam-webster.com/dictionary/bareback).
However, in the case of sex slang, it can also be used metonimically as an adjective or an adverb, meaning 'without a condom' or metaphorically, as a verb, denoting sexual intercourse with no protection (see: https://www.collinsdictionary.com/dictionary/english/barebacking).

beau palmarès de films à son actif, et exclusivement chez BelAmi où il tourne des vidéos ***bareback*** ('This young 26-year-old Czech already has a good list of films to his credit, and exclusively at BelAmi where he shoots bareback videos'); through the harder ones, e.g. (16) *Il est le genre de gars qui aime donner des claques durant ses ébats et aime la sodomie* ***hard***. ('He's the kind of guy who likes to slap during his antics and likes hard sodomy.'); (17) *Sans aucun tabou*, *il n'est pas rare de le voir dans des vidéos ultra* ***hard*** *et sans capotes* ('Without any taboos, it is not uncommon to see him in ultra hard and bareback videos') to the most extreme, e.g. (18) *l'as des plan abattages*, ***dog training*** *et autres délires fétichistes* ('the ace of slaughter sex encounters, dog training and other fetishist frenzies').

What is more, it is interesting to note here that the English adjective *extreme*, a borrowing through French from Latin, still exists in contemporary French; it has exactly the same meaning, though a slightly different spelling (*extrême*) and is perfectly variable in plural, which seems not to be the case in one of the following examples: (19) *On ne cesse de le redécouvrir dans de nouveaux* ***trips extreme*** *toujours plus excitant dans de nouvelles scènes porno* ('We keep rediscovering him in his new trips that are always more exciting in new porn scenes'); (20) *Très dominant et aimant les trips extreme* ('very dominant and loving extreme trips'). On the other hand, the word *hard* is also used in this type of a discourse even though, as Kacprzak puts it, this English adjective "is attested to by *Larousse*, but only as an element of compound borrowings *hard-core*, *hard-rock* and *hardware* where it occurs in the first position. The pattern followed by érotico-*hard* and *laïque hard* is different, because a French adjectival component is connected with an English adjectival component, but which has a value of an adverb" (Kacprzak, 2019, p. 98).

Besides the manifest denotations pertaining to various sexual practices, it is indispensable to highlight that other complete borrowings, which revolve around the semantic field of sex, refer to the actors themselves. More specifically, they may denote their sexual identity, e.g. (21) *Ce jeune mec* ***gay*** ('This young gay dude'); specific parts of their body, e.g. (22) *De son* ***bubble butt*** *avec lequel il adore jouer à ses pectoraux et ses abdominaux ciselés, Jordan adore montrer…* ('From his bubble butt he loves to play with, to his chiseled pecs and abs, Jordan loves to show off…'); their positions or, in other words, the roles they play as sexual partners, e.g. (23) *Avec sa bite d'acier, cet actif sait comment gérer les* ***power bottoms***. ('With his steel dick, this top knows how to handle power bottoms.'), or even tribes[11] to which they

11 In this socio-discursive landscape, *tribes* are regarded as "sub-communities of folks who identify largely based on their presentation" (see: `https://help.grindr.com/hc/en-us/articles/4402336949523-Building-your-profile-`), e.g. *Bear, Clean-cut, Daddy, Discreet, Geek, Jock, Leather, Otter, Poz, Rugged, Trans, Twink, Sober*. The aforesaid examples are derived from a classification of tribes offered by *Grindr*, known for being the biggest and most utilitarian gay dating app that is accessible to the public (see: https://www.gq-magazine.co.uk/article/best-gay-dating-apps). According to Levesley and Dawson (2023), "The most blessed and cursed thing about Grindr is – because it's so ubiquitous – that it really is a broad church represented by a wide range of members". Indeed, they are supposed to build their own profile by selecting a number of fields with a view to creating the most complete image of themselves, and therefore gaining the most attention. To do so, users

belong, e.g. (24) *Il cultive son look **geek**, avec ses lunettes qu'il quitte que très rarement durant les tournages* ('He cultivates his geeky look, with his glasses that he rarely takes off during filming'); (25) *Si vous aimez voir les petits minets se faire marteler le cul par un **daddy** dandy, regardez les vidéos de Dirk Caber* ('If you like seeing little twinks get their ass hammered by a dandy daddy, check out Dirk Caber's videos'); (26) *Rocco Steele est le **daddy** le plus chaud avec un sexe monstrueux. Il est l'un des **escorts** masculins les plus demandés de Manhattan* ('Rocco Steele is the hottest daddy with a monstrous cock. He is one of the most requested male escorts[12] in Manhattan'); (27) *Un regard clair, un style de **skater** et un petit cul* ('A clear look, a skater style and a small ass').

To sum up the above analysis of full borrowings, the first and the most prevalent type of lexical borrowings in our primary sources, we would like to make three accessory statements about the terms related to the semantic field of tribes[13] analyzed in the previous paragraph.

The word *daddy*, to begin with, which correlates closely with the term *son* in this socio-discursive realm, appears to have acquired a truly favorable connotation; according to Silverstein and Picano (2009, p. 78),

> Relationships of this kind respond to the real psychological needs of many gay men. More and more books about the important role of fathers in the sexual development of their sons have been published. As early as in 1981, Charles Silverstein, in his book *Man to Man: Gay Couples in America* described the desire some gay men feel for their fathers. (…) Clinical psychologists cite examples of teenage gay men who have tried to seduce their fathers. Richard Isay's book *Being*

---

of this application may edit their profile information and choose between the options under each of the following sections: *Basics* (display name and brief description of the user); *Stats* (age, height, weight, body type, position, ethnicity, tribes, relationship status); *Expectations* (types of connections, meeting places, and kinds of messages the user is open to receiving); *Identity* (gender identity, pronoun suggestions, including non-gendered pronoun options for languages which do not have gendered pronouns, e.g. Filipino, Hindi, Indonesian); *Sexual health* (several options concerning the user's health status); *Vaccinations* (self-reported vaccination status for Monkeypox, COVID-19, and Meningitis); *Social links* (sharing social media and personal information, e.g. Instagram, Spotify, Facebook); *Related articles* (settings, notifications, tags, language preferences) – see: `https://help.grindr.com/hc/en-us/articles/4402336949523-Building-your-profile-#h_01F-8DA858AENK628W33AYEAXJQ`

12 The French feminine noun *escorte*, a borrowing from Italian, has the same denotative meaning as the word *escort* in English; however, it is interesting to observe that the latter may also denote a person who is paid to go out socially with somebody, or, more euphemistically, somebody who works as a prostitute.

13 According to Maffesoli (2016), a sociologist who deals with the problems such as social ties in the community, the role of imagination, and the everyday life of today's societies, or, even more interestingly, tribes: "Between the 18th and the 20th centuries, modernity was built on rationalism, individualism (the individual seen as a unit) and on the social contract of the Republic, united and indivisible. The modernist project was to dominate nature and to be universal. The postmodernity which has succeeded it has seen the powerful return of the impulse to community and of the need for collective emotion – I call this *neo-tribalism* (1988)".

> *Homosexual: Gay Men and Their Development* discusses the importance of homosexual sons' erotic attachment to their fathers.

Then, it is the word *geek*, denoting someone fond of the new technologies, and particularly of computers, the Internet, video games[14], etc., that seems to be an enticing case of a complete borrowing from English in the French corpus. Not only can it be used only informally as both a noun and an adjective, but, even more interestingly, it also has its own grammatically feminine form (*geekette*), which is a notable example of a neological hybrid borrowing (English stem – *geek* and French feminine suffix – *ette*), and thus an epitome of commendable human ingenuity.

Finally, one ought to focus on the English word *skater* that is also far from being an uninteresting instance of a lexical borrowing in our corpus. Indeed, the French renowned dictionaries, such as *Larousse* or *le Robert* approve of using a slightly different version of this noun; in this case, it is another fine example of a hybrid borrowing, encompassing two grammatical genders: maculine (*skateur*) and feminine (*skateuse*)[15].

Clearly, the notion of tribes, so important in this type of discourse, is intimately connected with Le Breton's (2017) concept of the 'liberated' body – a conspicuous anthropological phenomenon that typifies the Western world of today. According to him,

> The body is 'liberated' in a fragmented way and cut off from everyday life. The discourse of liberation and the practices it gives rise to are indicative of middle or upper (privileged) social classes. This 'liberation' occurs less under the auspices of pleasure (even if, undeniably, pleasure is often present) than in the way of working on oneself – personalized calculation of which the matter is already given on the body market at a given time. This craze hardens the standards of physical appearance (to be thin, beautiful, tanned, fit, young, etc., for women; to be strong, tanned, dynamic, etc., for men) and maintains, more or less clearly, low self-esteem among those who do not produce the signs of the 'liberated body' (2017, p. 208).

In Table 1, we present the quantitative results of our investigation into full borrowings from English in the corpus studied.

Table 1. Full borrowings (14 units)

Examples of full borrowings	Number of occurrences
gay	39
sexy	7
bareback	6

14 https://dictionnaire.lerobert.com/definition/geek

15 https://dictionnaire.lerobert.com/definition/skateur

Examples of full borrowings	Number of occurrences
star	5
escort	5
hard	4
daddy	3
look	2
trip	2
geek	1
skater	1
bubble butt	1
dog training	1
power bottom	1

5.2. Hybrid borrowings

The second type of lexical borrowings that have been found in the course of the study are called hybrids. They are present in the explored discourse, but only on a very modest scale; more specifically, there are three different occurrences of hybrid borrowings, representing three different parts of speech: verb, noun, and adjective, yet referring to the same extralinguistic entity: X-actors as depicted in the erotic biographies that constitute the corpus we have explored for the purpose of this examination. In light of our analysis, it should be stated that a porn actor is flesh and blood, and therefore may well house a miscellaneous collection of human needs, emotions, desires, weaknesses, and so on.

On the one hand, he is perfectly able to be (28) *un excellent baiseur, viril, et diablement* ***performant*** ('an excellent fucker, virile, and devilishly successful'), the adjective *performant* being a hybrid borrowing (English word: *perform* and French ending: *-ant*) that denotes someone who is efficient, remarkable, competitive[16]. What is more, this favorably hyperbolic depiction of an X-actor is even strengthened by means of another striking case of a hybrid borrowing comprised of two linguistically foreign elements: English word (*hard*) and French suffix (*-eur*), typical of grammatically masculine nouns and adjectives[17]: (29) *Un* ***hardeur*** *hors pair ! Damien Crosse aime la baise bestiale et quand il déglingue…* ('An unequaled pornstar! Damien Crosse loves beastly fucking, and when he breaks down…').

16 https://dictionnaire.lerobert.com/definition/performant

17 https://www.larousse.fr/dictionnaires/francais/hardeur/10909968

In addition, it seems very interesting to notice that, rhetorically speaking, other examples containing the noun *hardeur* may well be interpreted as quite evocative figures of speech called auxeses. They are often used in persuasive discourses for emphasis or special effect: these laudatory hyperboles exaggerating unusual characteristics of X-actors are clearly indicative of the positive connotation carried by the word *hardeur*, as may be appreciated in the following quotations from our primary sources: (30) *Il a déjà tourné des scènes pour Lucas Ent. (bareback) et plus récemment en exclusivité pour CockyBoys où il partage les scènes avec des* ***hardeurs*** *confirmés* ('He has already shot scenes for Lucas Ent. (bareback), and more recently for CockyBoys exclusively where he shares the scenes with confirmed pornstars'); (31) *Avec des origines russes, ce* ***hardeur*** *ne cesse de nous surprendre par ses vidéos hard* ('This pornstar with Russian origins never ceases to surprise us with his hardcore videos').

On the other hand, it is worth noting that the language of pornography today has not been – as it is still quite often believed – fully manufactured by a broad scope of expletives. On the contrary, notwithstanding the thematic homogeneity of external pornographic discourse, it is far from being utterly homogenous in terms of style, register, tone, connotations, figurative language, etc., which can be illustrated by the word beginning with the letter f… *flirter*[18], another, much finer, example of a hybrid borrowing spotted in our corpus: (32) *Theo Ross transpire le beau mâle hétéro inaccessible, et pourtant, sous ses allures de mâle alpha il peut lui arriver de* ***flirter*** *avec d'autres mecs… et même se faire enculer !* ('Theo Ross is sweating the handsome, inaccessible straight male, and yet, beneath his alpha male looks, he can flirt with other guys… and even get fucked!'). Morphologically speaking, this informal verb also contains two foreign components: English word (*flirt*) and French ending (*-er*), characteristic of the first (and the largest) group of regular verbs.

In brief, it is crucial to conclude that all of the aforementioned borrowings from English play a considerable role in this socio-discursive landscape: they come to denote the human body, physical skills and carnal practices, which places the physique itself in the central position. Indeed, as Louis-Georges Tin put it, "tackling the gay issue with the question of the body means to recognize centrality of the body in gay culture" (2018, p. 234). In a broader perspective, however, not only is the anthropology of the body inherent in the LGBTQ community, but it also pertains to the modern world in a more general sense: "(…) the Western person today is animated by the feeling that their body is in some way other than them, that they possess it like a very special object, certainly more intimate than the others. The identity of substance between a human being and their bodily rooting turns out abstractly broken by this singular relationship of property: having a body" (Le Breton, 2017, p. 156).

In Table 2, we present the quantitative results of our investigation into hybrid loanwords from English in the corpus examined.

18 https://www.larousse.fr/dictionnaires/francais/flirter/34176

Table 2. Hybrid borrowings (3 units)

Examples of hybrid borrowings	Number of occurrences
hardeur	3
flirter	1
performant	1

5.3. Semantic calques

The third type of lexical borrowings – calques – distinguished during the study of our corpus is represented by one occurrence only: the adjective *versatile* seems to be a unique case in French erotic biographies: (33) *Souvent actif à l'écran, il en demeure pas moins* ***versatile*** ('Often top on the screen, he remains nonetheless versatile'); (34) *Kayden Gray a fait le grand saut en quittant la Pologne pour s'installer à Londres (…). Il est gay* ***versatile****, et aime en particulier la fellation…* ('Kayden Gray took the plunge and left Poland to settle in London (…). He is versatile and particularly likes fellatio…'). As unpretentious as it may appear, the word in question does constitute an outstanding example of a semantic calque.

Indeed, this positively-connoted adjective has been employed twice in the biographies in the same meaning as in the borrowing language ('able to do many different things'[19], which in the context of our analysis – one of the diastratic varieties of modern French – stands for [actors] 'able to play both top and bottom roles'[20]) although the act of borrowing it seems 'unnecessary' for two reasons. Firstly, it describes the extralinguistic reality for which the French language already has a semantically equivalent term (*polyvalent*)[21], used in its both standard and non-standard varieties. Secondly, the same adjective (*versatile*)[22], coming from Latin (*versatilis*, which literally means 'revolving')[23], is currently used in French, where it refers to someone who is fickle and changes their mind easily, and therefore carries rather disapproving connotations[24].

On balance, it should be emphasized that the specificity of the borrowings in our corpus is conditioned by the nature of the discourse examined – global, non-secret, external, and produced mainly in the English language. The human body, placed at the heart of this discourse, seems to be 'liberating' itself by fulfilling the need for anthropological roots alongside seeking for a new alliance with the carnal aspect of human identity. The latter, reinforced by an unprecedented advancement of new social media of the 21st century, has been crystallizing itself into a number of bodily practices of our time and testifying to the evolution of norms, tastes, preferences, etc.

19 https://www.oxfordlearnersdictionaries.com/definition/english/versatile?q=versatile
20 https://czasopisma.uni.lodz.pl/romanica/article/view/9855/9585
21 https://www.larousse.fr/dictionnaires/francais/polyvalent/62453
22 https://www.larousse.fr/dictionnaires/francais/versatile/81641
23 https://latin-dictionary.net/definition/38611/versatilis-versatilis-versatile
24 https://www.collinsdictionary.com/dictionary/french-english/versatile

(Paveau & Perea, 2014, p. 8). Hence, it is worth pointing out that "if the actor 'liberates' himself in these practices, it is not on his initiative; the atmosphere of a moment encourages him to do it according to certain modalities, but he engages himself into it with all the more personal commitment that he himself experiences the need to fight against the lack of being which comes from the underuse of his bodily energy. It is about achieving the fullest use of oneself, unifying the different levels of one's existence" (Le Breton, 2017, p. 187), as depicted, at least to some extent, in the erotic biographies scrutinized in this paper.

In Table 3, we present the quantitative results of our investigation into semantic calques from English in the French corpus.

Table 3. Semantic calques (1 unit)

Examples of semantic calques	Number of occurrences
versatile	2

6. Conclusions

On the whole, it is imperative to point out that borrowings constitute one of the commonest lexicogenic procedures in modern French. Indeed, focusing on neological adjectives in French, Kacprzak (2019) dedicates two sub-chapters of her book to exploring full adjectival borrowings as well as compound adjectives, and more specifically hybrid adjectival borrowings, from both classical (Greek, Latin) and modern languages (Arabic, English, German, Italian, Japanese, Spanish, etc.) According to her research, the greatest number of borrowings in both types of adjectives come from the English language: 68% of hybrid borrowings (2019, p. 99) and 77% of full borrowings (2019, p. 133).

Similarly, as a result of the analysis of the primary sources collected for the purpose of this study, a significant number of full borrowings from English was found in the diastratic variety of French examined in this paper. Secondly, a distinct advantage of this type of borrowings (14 units in total=ca. 77.8% of all lexical borrowings) over other types of lexical borrowings has been noted. Thirdly, among other types of loanwords occurring in the corpus studied, we should mention hybrid borrowings (three independent cases) and one case of calque – more specifically, an interesting example of semantic calque. Fourthly, it is worth emphasizing that no other types of calques or false borrowings have been recorded in the analyzed material.

The presence of borrowings in our primary sources (ca. 0.2% of all the words) seems important from the point of view of the most fundamental function that sociolects (in this case slang) can perform, i.e. the identity function (*sexy, star, look, gay, geek, daddy, escort, skater, hardeur, performant, label, fanbase*). The remaining functions of gay and bisexual slang appear less frequently in the corpus and are subordinated to the identity function: the ludic function (*bubble butt, trips extreme, sodomie hard, shows webcams, dodgeball, camping, go-go dancer*) and the

least frequently the ludic-cryptic function (*power bottom, versatile, top persistant, bareback, dog training*).

In effect, we are dealing here with external pornographic discourse, i.e. a social discourse of a generally non-secret nature, yet with a strong persuasive dimension. The main goal of the examined texts is to stimulate the imagination of recipients and, consequently, encourage them to use thematic websites and applications offering more and more possibilities. Ultimately, it is about creating an ethos based on identification; so is it in the widely studied hegemonic political discourse (Białas, 2021b, p.151) – the images that constitute it draw mainly from social emotion sprung from the shared values: "the citizen, through an irrational process of identification, melts one's identity into that of a politician" (Charaudeau, 2005, p. 105).

As regards the discourse explored in this paper, however, we find it interesting to emphasize the domination of North American porn studies over French scientific discourses related to pornography. Indeed, according to Landais (2014, p. 26), who delves into the case of American studies,

> although a certain number of *connections* are being taken into account by some researchers (Éric Maigret, Armand Mattelart, Érik Neveu) in several European countries and particularly in France, the human and social sciences are clearly lagging behind North American studies. Nevertheless, information and communication sciences or sociology are increasingly asserting themselves in connection with the latter, and hence are creating a common paradigm. In addition to being interested in the media 'because they are the large area of the production of identities and powers (Maigret 2013: 160), they favor the approach of communication and tackle the subjects which, until now, have been considered illegitimate by the scientific community as dealing with the *dirty outside world*.

In Table 4, we present the overall quantitative results of our investigation into lexical borrowings from English in the French corpus.

Table 4. Total number of lexical borrowings from English in the corpus studied

Full borrowings	14	77.8%
Hybrid borrowings	3	16.7%
Semantic calques	1	5.6%

References

Primary sources

Erotic biographies accessed from: https://www.videosxgays.com/. (April 2024)

Secondary sources

Białas, M. (2021a). L'amour dans toute sa nudité: le langage des biographies d'acteurs X gay et bisexuels sur les sites Internet pornographiques. *Folia Litteraria Romanica*, *16*, 141–148.

Białas, M. (2021b). Entre pornographie et politique. La fabrication de l'ethos de puissance dans le discours pornographique externe: le cas des biographies érotiques d'acteurs X homo- et bisexuels en ligne. *e-Scripta Romanica*, *9*, 148–157.

Le Breton, D. (2017). *Anthropologie du corps et modernité*. PUF.

Charaudeau, P. (2005). *Le discours politique. Les masques du pouvoir*. Vuibert.

Goudaillier, J.-P. (2001). *Comment tu tchatches! Dictionnaire du français contemporain des cités*. Maisonneuve & Larose.

Kacprzak, A. (2019). *La néologie de l'adjectif en français actuel*. Wydawnictwo Uniwersytetu Łódzkiego.

Kozłowska, M. (2021). *Les anglicismes dans la presse féminine française: étude du langage de la mode sur la base des textes sélectionnés de l'édition française du magazine Elle (2018 – 2020)*. [Unpublished bachelor's thesis]. University of Białystok.

Landais, É. (2014). *Porn Studies* et Études de la pornographie en sciences humaines et sociales. *Questions de communication*, *26*, 17–37.

Loubier, C. (2011). *De l'usage de l'emprunt linguistique*. Office québécois de la langue française.

Maffesoli, M. (2016). From society to tribal communities. *The Sociological Review, 4* (64), 739–747.

Paveau, M.-A. (2014). *Le discours pornographique*. La Musardine.

Paveau, M.-A. & Perea, F. (2014). *Un objet de discours pour les études pornographiques. Questions de communication*, *26*, 7–15.

Picano, F. & Silverstein, C. (Eds.). (2009). *Radość seksu gejowskiego*. Wydawnictwo Czarna Owca.

Tin, L-G. (2018). Gai. In Andrieu, B. & G. Boëtsch (Eds.), *Dictionnaire du corps* (pp. 233–237). CNRS Éditions.

Internet Sources (accessed in April and May 2024)

https://www.collinsdictionary.com

https://dictionnaire.lerobert.com

https://www.larousse.fr

https://latin-dictionary.net/

https://www.macmillandictionary.com

https://www.merriam-webster.com

https://www.oxfordlearnersdictionaries.com

https://www.thesaurus.com
https://www.academie-francaise.fr/
https://czasopisma.uni.lodz.pl/romanica/
https://www.gq-magazine.co.uk/
https://help.grindr.com/hc/en-us

***

Mateusz Białas holds a Ph.D. in Theoretical, Descriptive, and Automatic Linguistics from Université Paris Cité and in General Linguistics from the University of Warsaw. He is a conference interpreter, a certified English and French teacher, and an assistant professor at the University of Warsaw (Poland). In the years 2018-2024, he worked in the Department of Lexicology and Pragmalinguistics at the University of Białystok, Poland. He specializes in the sociolinguistics of language practices, focusing particularly on persuasive discourses such as hegemonic political discourse and external pornographic discourse. His research interests and activities are predominantly centered on rhetoric, textual and discursive linguistics, pragmatics, lexical semantics, direct and relay simultaneous interpreting, as well as the study of emotion, identity, and body within the language sciences.

LARYSA CHYBIS[1]
Curtin University, Perth, Western Australia
https://orcid.org/0000-0003-2969-239X
SALLY LAMPING
Curtin University, Perth, Western Australia
https://orcid.org/0000-0002-9496-039X
TONI DOBINSON
Curtin University, Perth, Western Australia
https://orcid.org/0000-0003-1790-0016
KATHRYNE FORD
Curtin University, Perth, Western Australia
https://orcid.org/0000-0001-7685-2953

DOI: 10.15290/CR.2024.45.2.03

English as a barrier on the pathway of professional transitioning of Ukrainian migrant teachers in Australia

Abstract. Australia is seen as a promised land by migrants pursuing a better future for themselves and their families. Highly qualified migrants with vast work experience frequently encounter hurdles on their way to professional realisation, including the official language of the country – English. This study investigates the difficulties Ukrainian teachers face with the English language on their professional transition pathway in Australia. The research involves Ukrainian migrant teachers who obtained a specialist or master's degree in Ukraine or another post-Soviet country, whose professional experience in Ukraine was in the teaching field, and who identify as Ukrainians. Narrative interviews, memos, documents, and artefacts are the data collection methods; thematic analysis is used to unpack the data. The participants were chosen using non-probability purposive snowball sampling, which engages contacts within the community and benefits projects with a small number of respondents. The inquiry elucidates the extent of the English problem for Ukrainian migrant teachers on their way to professional transition in Australia. The research will benefit other migrant teachers facing similar barriers when trying to re-enter their profession in the new environment and inform apposite institutions about the existing hurdles to facilitate positive changes in the field.

1 Corresponding author: School of Education, Curtin University, Perth, Western Australia. E-mail: larysa.chybis@postgrad.curtin.edu.au

Keywords: professional transition, Ukrainian migrant teachers, the English language, professional identity, narrative study

1. Introduction

English-speaking countries, including Australia, have welcomed and provided opportunities for residency and citizenship to migrants (Dinesen et al., 2020). Australia promotes the ethos of multiculturalism (Marcus, 2014) by implementing different migration programmes, thereby enriching the Australian population with individuals from a diverse national, ethnic, cultural, religious, and linguistic background (Wang et al., 2023). However, certain English varieties enjoy more privilege than others (Tupas, 2021), despite substantial research on the structural, sociolinguistic, and political legitimacy of various English varieties worldwide. Employees identified as native English speakers retain a privileged position in the workplace (Harrison, 2013). From the native speakerism perspective, speaking English as their first language automatically makes a person the best teacher of English (Keaney, 2016) – this has prevailed in the field of teaching English for decades. However, Gilmour et al. (2018) assert that non-native English-speaking teachers (NNESTs) have certain valuable assets in reserve that are deficient in predominantly monolingual native English-speaking teachers (NESTs). The UNESCO database shows that it is more effective to use a method referred to as "multilingual education" when teaching English as a second language (UNESCO, 2023). For decades, Ukrainians have been immigrating to Australia. The ongoing war in Ukraine has prompted a fresh influx of Ukrainian migrants inclusive of teachers (Department of Home Affairs, 2022). Ukrainians rank among the most sizable ethnic groups residing outside their homeland's borders (Fedyuk & Kindler, 2016). Nevertheless, no research regarding Ukrainian migrant teachers has been conducted yet.

This article aims to explore linguistic aspects of the professional transition of Ukrainian migrant teachers in Australia by attempting to answer the following research questions:

— What linguistic challenges do Ukrainian migrant teachers encounter in transitioning to employment aligned with their skills and training obtained overseas?
— How do Ukrainian migrant teachers overcome these obstacles on their professional pathways in Australia?

2. English language hierarchy

The English language has undergone a transformation from a colonial language to a lingua franca spoken by people throughout the world. Frequently, English is associated with education and culture and has become a language of communication in academia (Lai, 2021). A quarter of the global population speaks English (Ibrahim et al., 2019), and it has become the best option for people speaking other languages throughout the world (Jenkins, 2019).

There are several accepted or native Englishes, such as British, North American and Australian. The market value of different forms of English varies, and not all Englishes are considered

equal. In Australia, British English holds a higher status compared to American English (Harrison, 2013). These Englishes have their own pronunciation, dialect peculiarities, and hierarchy, with Standard English and Received Pronunciation residing at the apex of the linguistic pyramid (Trudgill, 2002). Conversely, Englishes acquired in countries where the national language is other than English are at the bottom of the hierarchical pyramid.

Given the global influence of English, one's English fluency is a means of exercising power: Standard English is often linked to status and authority. Proficiency in Standard English can play a crucial role in unlocking opportunities for a person's future endeavours (Brady, 2015) and allows access to the huge database of human knowledge worldwide (Zeng et al., 2023). Individuals who do not possess this standard language may experience a sense of disempowerment. The term "non-standard" implies notions of being less valid in some respects (Baratta, 2018, p. 59).

In Australia, the English language holds significant value as a prominent linguistic asset due to its prevalence in education, employment, and economic activities. Its elevated status perpetuates the societal structure along with specific standards of the language (Harrison, 2013). Australia, as a dominant language nation, gravitates towards monolingualism, where being monolingual is a choice rather than a lack of opportunities (Lai, 2021). However, learning the dominant language is not an option but a necessity for people migrating from linguistically diverse backgrounds. Bilingualism is a potential advantage that improves individuals' employment prospects, though the language abilities of bilingual speakers are often deemed deficient when assessed against monolingual standards (Harrison, 2013).

3. Translingual discrimination

Translingual discrimination refers to the phenomenon rooted in the unequal dynamics between transnational migrants and the dominant population of the host country, primarily centring on language as a pivotal element in this interaction (Dovchin, 2022). Native speakers from English-speaking countries often maintain the belief in the superiority of their own form of English (Harrison, 2013). This conviction may lead them to feel entitled to assess and judge the English of other users. Dovchin (2022) distinguishes the following types of translingual discrimination: translingual name discrimination and translingual English discrimination. The former occurs when individuals with names associated with a dominant culture are stereotypically perceived to possess superior skills compared to those with names reflecting a transnational background. Hence, ethnic minorities may adopt names typical for the dominant group out of fearing discrimination based on their original names (Kang et al., 2016). Members of the majority group tend to view individuals from ethnic minorities who choose Anglophone names as displaying a greater commitment to assimilation, suggesting a closer alignment with cultural beliefs and values compared to those who retain their original names (Zhao & Biernat, 2017). Moreover, original names can pose challenges in terms of pronunciation. Existing research indicates that individuals with names that are hard to pronounce may face less favourable judgments (Laham et al., 2012). Translingual English discrimination refers to the exclusion or ostracism of

transnational migrants based on their English language proficiency and pronunciation, and it can be bifurcated into accentism and stereotyping (Dovchin, 2022). Although there is no voice devoid of an accent (Sterne, 2022), some so-called native accents are preferred over non-native English accents. Having a distinct accent compared to Standard Australian English and lacking familiarity with local expressions and customs signal that someone may not belong to the group (Tankosić, 2022). Thus, being different from the dominant group elicits negative sentiments from the locals holding hegemonic ideologies even towards a migrant speaking English fluently and clearly (Dryden & Dovchin, 2021). In the communication between native and non-native speakers, the latter may feel anxious about receiving unfavourable judgments based on their speech, prompting the person to strive to improve their communication (Birney et al., 2020). Paradoxically, the effort invested in this response, coupled with underlying anxiety, could impede effective communication in English. According to Aichhorn and Puck (2017), anxiety poses a major problem for non-native speakers of English, which can provoke nightmares, sleeping disorders and self-doubt. Physically, it can manifest in hyperhidrosis (excessively sweaty hands), excessive perspiration, and a sensation of shakiness or agitation.

The coping strategy involves establishing a "safe translingual space" for migrants, providing them with a supportive environment where they can gather, communicate, and share their concerns without constantly worrying about conforming to the standardised requirements of English speech (Dovchin, 2022, p. 56). Kiramba and Harris (2019) emphasise the importance of creating interactive environments where bilingual or multilingual individuals can delve into their metalinguistic skills and showcase their multilingual competencies.

To avoid discrimination based on their accent, some enrol in classes with the goal of eliminating their accent, expressing a desire to achieve a fully native sound in their speech (Oleinikova, 2020, p. 229). These migrants, identified by Oleinikova (2020) as "achievers", distance themselves from their native language and culture and avoid contact with representatives of their ethnic group as a potential hindrance to their career growth. This leads to ethnic evasion of these migrants and partial loss of their identity as one's accent serves as a phonological indicator of an individual's identity (Rangan et al., 2023) and, modifying one's accent, "you lose part of yourself" (Park & Lee, 2022, p. 34).

4. Non-native English-speaking teachers

Despite numerous efforts made in recent years to address the ostracism of NNESTs in the language teaching profession, the English language teaching field continues to exhibit a contrary trend. The teaching environment remains structured and defined by boundaries influenced by language ideologies, resulting in the exclusion of NNESTs from participating in foreign language teaching on an equal footing with NESTs (Wright, 2022).

NESTs are frequently perceived as superior and credible, while NNESTs may be unjustly portrayed as inferior and incompetent, sometimes to the extent of being labelled as an underclass (Rivers, 2018). This bias is evident in the job market, where native English speakers are often

in higher demand and receive better compensation, even with only a short course certificate in English language teaching among their qualifications. Global job postings explicitly seeking English teachers exclusively from Great Britain, the United States of America, Canada, Australia, and New Zealand further support this trend (Wright, 2022). Thus, non-native English teachers frequently grapple with feelings of shame and develop inferiority complexes, comparing themselves to their native counterparts (Rivers, 2018).

NESTs have long been in demand in the field of teaching English. From the native speakerism point of view that has dominated English Language Teaching (ELT) (Ershadi, 2024; Holiday, 2005), NESTs have been considered to be the standard to follow with regard to the language, which automatically made NESTs the best teachers of English including the methods of teaching. However, this notion has been increasingly criticised from the perspective of its racism, fallacy, linguistic imperialism, prejudice and discrimination. Some NESTs do not have sufficient pedagogical experience or qualifications and are predominantly monolingual (Keaney, 2016).

According to Phillipson (2012), the ELT myth that the best teacher is a native speaker supports linguistic imperialism. Many NESTs may lack adequate understanding or background in the specific teaching practices needed to support learners of English as a second language (Oliver et al., 2017). UNESCO (2023) has stressed the importance of native languages and multilingual education. Phillipson (2012) asserts that the positive effect of multilingual education based on the person's native language is definite, whereas English-based education ignoring native languages and cultures is inadequate and unproductive. NNESTs possess professional assets that are often absent in their monolingual colleagues (Gilmour et al., 2018), who comprise the majority of Australian schools. These assets include being multicultural and multilingual, having versatile pedagogical knowledge, and having valuable teaching experience (Marom, 2019).

5. Ukrainian migrant teachers' professional transition and identity work

An individual's identity is shaped through interactions with others and the wider socio-cultural environment (Maydell, 2020). Accordingly, self-perception is socially formed and linked to distinct social roles (Stets & Serpe, 2013), often referring to social categories, namely occupation, ethnic group, or marital status, when introducing themselves (Schwartz et al., 2011).

Identity construction and reconstruction are continual processes for everyone, but migrants undergo more pronounced changes (Maydell, 2020). This can entail creating a new identity or adapting their existing one to the new environment or circumstances. The identity crises faced by refugees, displaced persons, and migrants necessitate a transformative shift in their identities (Alieksieienko, 2020). This transformation is essential for developing effective strategies that facilitate their seamless alignment into Australian society (Oleinikova, 2020). The identity work process relies on the addition, retention and subtraction of elements within one's identity (Lepisto et al., 2015). Adding to a person's identity constitutes either embracing a new identity or appending features to the existing one to better align with the new setting (Ennerberg &

Economou, 2021). Retaining an individual's identity entails maintaining and fortifying one's sense of self, whereas subtracting identity leads to a partial or complete loss of one's identity (Lepisto et al., 2015).

6. Materials and methods

6.1. Methodology

The focus of this research is the professional transition of Ukrainian migrant teachers in Australia. The narrative research design entails comprehending lived experiences from the viewpoint of the respondents rather than striving to portray life objectively (Mertova & Webster, 2019). The research is grounded in an interpretivist paradigm, specifically well-suited for narrative studies (Papathomas, 2017), as its objective is to explore the complexities of the world by studying personal experiences within a societal context (Schwandt et al., 2007). The concepts of identity and narrative are often theorised as intertwined (De Fina, 2015), as an individual's identity tends to be expressed and molded through the act of narrating their story. Additionally, a narrative study has been chosen for its focus on the nuanced and complex facets of human experiences, particularly their relevance to the personal experience stories of teachers (Creswell, 2019). The comprehensive nature of the research design enables researchers to encompass the entire narrative based on the principles of continuity across the participant's past, present, and future (Clandinin, 2016). This stands in contrast to other research methods concentrating on specific moments, which potentially overlook noteworthy nuances interwoven throughout the enveloping storyline (Mertova & Webster, 2019).

The research has been conducted by the PhD student, Ms Chybis, referred hereto as "project lead", supervised by a team of experienced researchers of Curtin University located in Perth (Western Australia), further referred to as "research team" for convenience.

The credibility of this research, a criterion of qualitative research trustworthiness (Korstjens & Moser, 2018), has been achieved by prolonged engagement, persistent observation, triangulation and member check. The research entailed prolonged engagement with the participants by conducting three interview sessions to narrate their stories, clear up the collected data and ask follow-up questions. The participants were encouraged to bring memory boxes that helped them tailor their narratives and highlight certain intricacies and heartfelt moments of their lives.

The project lead has observed the principle of persistent observation by reading and re-reading the collected data and then working with the research team to negotiate the coding process for the data and the resulting thematic analysis. The memos written by the project lead during the interviews provided depth to the narratives and helped the research team reflect on the collected data which were later coded. Themes were identified before conducting the final set of interviews. Data triangulation has been achieved by using multiple methods of data collection, such as narrative interviews, documents, memos and artefacts. After a discussion with the research team, the project lead undertook a negotiated process with the research participants

in order to address any potential discrepancies between the original and restoried narratives (Creswell, 2019) and has been discussing the data codes and themes with the research team to check the inquirer's interpretations.

6.2. Data collection methods

The primary method of collecting data necessitates narrative interviews because this interview type enables respondents to casually relate their stories, diverging from the conventional question-and-answer format (Kartch, 2017). The interviews were planned in consultation with the research team and comprised three sessions, with the initial two typically taking place on the same day and lasting up to 90 minutes. In the first session, the project lead listened to the participant's narrative without interruptions. Subsequently, questions were posed to the participants to seek clarification on specific aspects related by the interviewees during their story (Burke, 2014). Afterwards, the project lead discussed these interviews with the research team, and it was decided that additional interviews would be scheduled with the respondent to authenticate and substantiate the collected data.

Narrative interviews were complemented by documents, artefacts and memos to provide a comprehensive depiction of personal stories. The project lead took copious notes in order to elucidate and traverse data, as well as to determine new ideas and motifs. Writing memos during the interviews enabled the documentation of observations pertaining to interviewees' non-verbal communication and emotions.

The interviewees were encouraged to present their 'memory boxes' with personal artefacts at the first interview sessions. This was done to elicit otherwise forgotten memories, to enrich their personal narratives, and to navigate the process of narrating their personal stories.

6.3. Positionality

The incentive for this research originates from the project lead's personal experiences as a Ukrainian migrant teacher in Australia. The project lead acquired a fervid curiosity about Ukrainian migrant teachers' personal experiences in Australia after contending with a dearth of connections within the Australian teaching environment and the difficulty of adjusting to the local educational system when investigating her employment perspectives that would match her expertise obtained in Ukraine. The methodology for the research has been chosen given the project lead's Ukrainian background and close connections with the Ukrainian community. The project lead has managed to secure prospective respondents for the research, being a member of the Ukrainian community in Western Australia and a volunteer English teacher for displaced Ukrainians. The project lead and the participants have developed a communal affinity due to their shared linguistic background, ethnicity, and culture. Being a Ukrainian migrant teacher and an internally displaced person have enabled the project lead to examine the issue with the depth of an insider (Beresniova, 2017) and to provide a comprehensive study of its intricacies that monolingual scholars without migrant experiences may find daunting to fully grasp (Radhouane et al., 2022).

The project lead acknowledges the potential influence of her background, experiences and personal interactions with the participants in the project. Thus, the field texts and coded data have been routinely reviewed by the research team to eschew the project lead's biases, which could influence the recognition of themes (Patton, 2015). During the follow-up interview sessions, the project lead sought to clarify information provided by the interviewees, which seemed to be ambiguous. The data collected during the first interview sessions were coded before conducting the final sessions with the participants in order to differentiate between the project lead's and the participants' experiences. Additionally, the project lead has prolifically used quotes from the interviews to allow an inclusive and diverse representation of the participants' voices.

6.4. Participants and setting

The main focus has been on studying complete stories since the research design utilises a narrative methodological approach. The number of research participants has been restricted to five, given the encompassing nature of their individual experiences and the project lead's need to meticulously record the nuances of these stories. A smaller participant pool has enabled the project lead to preserve the data complexity (Creswell, 2019) and allowed her to commit sufficient attention to comprehensive analysis and subsequent data collection.

The project lead has employed a non-probability purposive snowball sampling method utilising her connections within the Ukrainian community in Western Australia to identify potential participants for the project. This approach is especially well-suited for endeavours with a restricted pool of participants (Cohen et al., 2018). The prerequisites for selecting participants were the following: participants must affirm their Ukrainian identity; their primary employment in Ukraine or another country in the post-Soviet space was teaching in a secondary or tertiary educational institution; and they must possess either a master's or specialist degree (master's equivalent in accordance with Legislation of Ukraine (2014)) obtained either in Ukraine or another post-Soviet country.

6.5. Data instruments and analysis

The collected data have been processed following an inductive approach, which includes the generation, analysis, and reporting of themes. In this approach, the coding of data was devoid of any prior theoretical framework or predefined ideas. Rather, the process of coding was developed involving the themes drawn from the interview data. This method of processing data enabled a versatile and unrestricted examination of the participants' stories, avoiding the imposition of predetermined categories or theories during analysis.

The analysis of qualitative data has encompassed the restorying process. First, the respondents' personal narratives in the form of field texts or raw data (Clandinin, 2016) were collected during the narrative interviews. Then, the interview recordings were transcribed and, if it was required, translated from Ukrainian into English. Afterwards, the data underwent thorough analysis for the key elements: characters, settings, problems, actions, and resolution (Creswell,

2019). Recurring themes regarding the experiences of Ukrainian migrant educators in Australia were ascertained while reconstructing the narratives. In a subsequent stage, the individual narratives of the respondents were woven together into a larger overarching tale, emphasising significant themes rather than existing as separate accounts.

6.6. Ethical considerations

The Human Research Ethics Committee of Curtin University approved the project, ID: HRE2023-0115. The project lead provided the participants with the project information, including the consent forms in either English or Ukrainian, depending on the participant's English fluency. Likewise, the interviews were conducted in Ukrainian or English, depending on the interviewee's choice. Additionally, the project lead explained the research objectives to the participants before conducting the interviews. The project lead disclosed the voluntary nature of the research project to the participants, who were free to withdraw from it at any time or decline a question during the interviews if it made them feel uncomfortable. The project lead has concealed the participants' identities and personal information that could lead to the participants' identification.

The project lead, being a representative of the same ethnic group as the participants and being a migrant herself, availed her experiences and empathy to create an environment comfortable for the interviewees to share their individual stories. Also, the project lead has obtained a Standard Mental Health First Aid certificate in order to identify distress and trauma symptoms and to readily inform the participants of the professional counselling opportunities in Western Australia.

7. Findings and discussion

The findings outlined below are part of the research conducted to obtain a Doctor of Philosophy in Education. This study has delved into the professional transition of migrant teachers who moved to Australia from Ukraine and one post-Soviet country and who identify themselves as Ukrainians, with Ukrainian being their native language.

7.1. Participants' overview

Four of the Ukrainian Migrant Teachers (UMTs) are specialists in English, as presented in Table 1; three of them were trained in Ukraine and one UMT obtained her English teacher qualification in another post-Soviet country. One UMT received training as a mathematics teacher at a Ukrainian university. All the UMTs are females aged between 32 and 70, with their pedagogical experience ranging from two to 48 years. Three of the UMTs arrived in Australia holding their student visas, and two fled Ukraine when the large-scale stage of the war started in 2022. Four out of five UMTs continued their professional education in Australia. The time spent in Australia varied from over a year to 10 years during the first interview session. Four UMTs stayed in the pedagogical field after their arrival in Australia, and one chose another field for developing her career.

Table 1. UMTs' gender, age, grounds for coming to Australia, duration in Australia, pedagogical field, teaching experience overseas, jobs in Ukraine and Australia

UMTs	Gender	Age	Grounds for coming to Australia	Duration in Australia	Pedagogical field	Job in Ukraine	Teaching experience overseas	Job in Australia
UMT1	Female	37	Student visa (for PhD degree)	10 years	English	Linguist	Eight years	Academic
UMT2	Female	32	Student visa (for Master's degree)	Five years	Italian and English	Language schools, university lecturer	Five years	Hotel business, government job
UMT3	Female	70	War	Over a year	English	Secondary school teacher of English	48 years	Teacher assistant
UMT4	Female	37	Student visa (for PhD degree)	Seven years	English	University tutor	About seven--eight years	Relief teacher, university tutor
UMT5	Female	39	War	One year and a half	Mathematics	Secondary school teacher of mathematic	Two years	Teacher assistant for disabled children

7.2. English fluency

English fluency can pose a significant hindrance for migrant professionals on the way to their professional transition in an Anglophone country. UMT5 had to seek employment at an educational institution other than a mainstream high school because of her insufficient English fluency: "They said that taking into account my language barrier, it would be more difficult for me, in general, to find a job in a mainstream school". Four UMTs, being teachers of English by their training and occupation, were fluent in English. Three of them, who had student visas on their arrival in Australia, passed their International English Language Testing System (IELTS) academic test as their visa requirement, which can serve as another indication of a person's fluency in English. UMT2 said that being a teacher of English gave her the speaking skills necessary in the English-speaking country: "Having a teaching background … was very [helpful]. I didn't have any language barriers, so it was not something that I was worried about… It was very easy for me". She added that her English fluency was sufficient for a career in Australia: "My English was definitely enough for all of the jobs that I was doing".

UMT5 received her pedagogical training as a mathematics teacher. However, English had never been an issue for UMT5 as she never planned to leave Ukraine and used the language as a means of communication when travelling abroad:

> I have never been concerned about my English. It was enough for me to be able to go somewhere in Europe as a tourist in order to converse with the denizens. […] I never thought that I would need to know English so well to be able to go to a school and teach there.

Nevertheless, English became quite a challenge when UMT5 had to flee the war in Ukraine in 2022: "It became a real problem when I migrated here". UMT5 described her English fluency among locals soon after she arrived in Australia as being "numb": "First, I was sitting at the table as if I was numb". However, her English fluency was assessed high enough as UMT5 was able to join the class for Certificate III of the Adult Migrant English Program (AMEP), which helped her to improve her English:

> I am really grateful to the Australian Government for these programmes in TAFE [Technical and Further Education], which enable one to improve their English to the point when one can communicate with locals and feel a bit better in society.

7.3. Different Englishes in Australia

Some of the UMTs stated that English in Australia differs from English varieties spoken elsewhere. UMT5 compared Australian English to the one spoken in Europe: "Australians speak English which is different to the one spoken in Europe. All the same, it [English] is different. That is why it was very difficult for me". She admitted that in the beginning, comprehending the Australian version of English posed such a challenge to her that she felt as if she was "deaf and numb":

> If in Europe I could communicate with people to some extent, in Australia, there is an absolutely different accent, and people speak somewhat differently here. I remember feeling as if I were deaf and numb for the first few months. You could say nothing to people, and you could not understand what people were trying to tell you.

One of the reasons for such communication problems is different Englishes and diverse accents that exist within Australia because of the multicultural nature of the country:

> I have been communicating with migrants, for example, in transport when I needed to get somewhere, and I asked people. They were, for example, people with dark skin, so I suspect they were not Australians. They might be natives of Africa. They were also people who had a specifically Asian appearance. They were speaking differently, too.

Similarly, UMT2 admitted that the diversity of accents in Australia was a problem for her to some extent, requiring some adjustment:

> Different accents, that was a problem. [The] Indian accent. I [had] never seen Indians before I came to Australia, and it was very hard for me to understand them in the beginning. As soon as you get used to their accent, … you … understand everything.

UMT3 described the local Australian accent as "challenging": "It [Australian English] is a bit challenging for me". She found other migrants' accents easier to comprehend compared to the local Australian accent: "In terms of the languages, yeah, sometimes, it's easier to understand people of other nationalities". She described it as sharing diverse cultural backgrounds or being in "the same boat" in her own words: "Maybe, it's easier to understand migrants because, for migrants, English is also a second language. We're just in the same boat with them, and I just understand them".

UMT4 asserted that she had no issues regarding academic English as she came to Australia for her PhD course: "I had no problem in terms of academic English, and at university, I was pretty good with communication and everything". Nevertheless, everyday Australian English became a communication barrier: "When I was attending some social gatherings, I had trouble understanding slang. I had trouble sometimes understanding the accent". UMT4 recalled her first linguistic crisis in Australia: "My first problem was when I came to Australia and tried to order some […] fast food. I didn't understand a word". Despite her degree in TESOL and a good understanding of academic English, this academic experienced a communication dysfunction when talking to a fast food employee speaking with a broad Australian accent.

UMT2 has also mentioned that Ukrainian teachers of English are accustomed to Standard English, to which Ukrainian educational institutions adhere. In her view, it becomes a problem when Ukrainian professionals migrate to Australia:

> People come [here] even with a teaching background; they are professionals, and they know English well. In Ukraine, we were not studying real English. We were learning perfect British English that no one else in the world is probably speaking.

UMT1 was accustomed to academic English but faced problems with everyday communication. She differentiated between academic English used in Australian educational institutions and everyday English requiring different vocabulary knowledge and fluent speaking skills:

> …even though I worked with Australians, it wasn't your everyday life. While here, I found that when I lived on my own…, I would be working a lot doing my PhD on my own, which wasn't very helpful in terms of everyday… just training your brain to, I guess… especially in terms of oral speech, of presentation, all of those phrases. It was much harder to speak than to write or to read.

UMT1 solved her problem with everyday English by exposing herself to Australian English so she could get used to the dialect:

> I was visiting friends, so I spent some time with them, talking to them from morning till evening. I realised that my brain gets that stimulation, and it becomes much easier for me after that. I think that was something that helped.

UMT3 partially distanced herself from locals who could be "talking too fast", giving her preference to communication with other educated migrants and Australian teachers like herself. Sharing the same professional field and Australian teachers' clear English contributed to the comfort of their communication:

> In terms of the languages, sometimes, it's easier to understand people of other nationalities maybe, but sometimes […] it depends on their education. Some people, even Australians, […] they're talking too fast, and it's difficult for me just to understand them, but some Australians are just like teachers; for example, their language is very easy for understanding, and I find it much easier to communicate with these people than with locals… no problems communicating with teachers. […] but when it comes to the locals, […], like any other professions, or especially old people, sometimes, I don't quite understand what they mean.

7.4. Career in Australia

Four of five UMTs have remained in the teaching field after moving to Australia. UMT2, who was also a teacher of Italian and English in Ukraine, has chosen to develop her career in a different field, doubting herself as a teacher of English in the Anglophone country: "I would not consider teaching English in an English-speaking country". During the interview, UMT4 asserted that

she experienced self-doubt as an NNEST in Australia: “I felt that I’m not good enough to teach somebody English”.

Moreover, UMT4 expressed her irritation regarding prejudicial attitudes towards NESTs and NNESTs that exist in society, including how it has resulted in unfair treatment of NNESTs and competition between these categories of teachers that she experienced when working overseas:

> I was fed up with the treatment because when you teach English as a foreign language, and you have tutors who are native English speakers, you always have to compete with them and prove that your knowledge is good enough. Even though the native English speakers are not really trained to be teachers and teach English as a foreign language because they just speak the language as native speakers, and they never learned pedagogy behind what they teach.

Being an NNEST in Australia, UMT4 experienced difficulties securing a job that would align with her training and experience: “I was trying to look for the jobs like English tutor and being a non-native speaker made it quite difficult”. UMT1 confirmed that it was difficult for her to find a permanent job, which was an additional stress: “Not knowing whether I’ll have an employment next semester was quite stressful”. UMT5 was not able to find a job in a mainstream high school in Australia that would match her pedagogical training and teaching skills being a non-native English speaker: “I wanted to work in high school… because I thought that being a teacher of mathematics I would make a greater contribution in high school”.

The UMTs encountered other challenges on their pathway to professional realisation in Australia: their Ukrainian accent and non-Anglophone names. UMT4 received numerous rejections from potential employers partly because of her name, which revealed her otherness. The UMT had to change her name in order to start receiving responses to her job applications:

> My name in my CV sounded like the person they don’t want to work with. I asked career consultants to help me with my CV, and the first thing they asked me to do [was] to change my name so that it sounded easier for people to read it.

Cases of accentism were reported during the interviews, such as in UMT4’s story when the native-speaking person testing migrants’ English speaking skills showed their prejudice: “I definitely felt that the examiner was biased and was criticising me for my accent. The facial expression wasn’t friendly at all”. UMT4 felt vulnerable regarding her Ukrainian accent when required to speak English: “Every time when I start talking, people ask me, “Where are you from?” which indicates that my accent gives me away”. For UMT4, her accent was something she might not like but had to put up with: “I came across the fact that we have the accent, and I had to live [with] that fact”.

UMT5 admitted experiencing difficulty being accepted as a team member and feeling like she was an outsider among local counterparts because her English fluency was insufficient for

leading everyday conversations and understanding people. She said that she was not "accepted" by her local colleagues because she was a "stranger" to them:

> I have noticed that Australians… If a total stranger comes, it is very difficult for this person to be accepted as a team member. Especially if there is a language barrier… You cannot understand people well or express your thoughts. Therefore, you are a bit isolated all the time.

Another problem that the UMTs encountered and felt was unfair was the necessity to take the IELTS test. UMT5 referred to the IELTS test as the "main problem". Some of the UMTs had to take the test on numerous occasions in order to prolong their stay in Australia and to be able to find a job in their professional field:

> Because it expires every two or three years… that can be quite stressful because it is quite a long test. It is also very expensive, and just understanding that once you reach the band you need, and you feel like, "Okay, I've done that," I think that is a stage of my life in the past, then you have to do it again. You have to prove yourself every time (UMT1).

UMT1 describes her experience of repeatedly having to take IELTS as "stressful", "expensive" and something that deprived her of the sense of stability and made her future seem uncertain. UMT4 had to take the test twice and admitted feeling similar sentiments: "I was really anxious about taking English tests".

Volunteering in the local Ukrainian community school enabled some of the UMTs to resume their teaching roles and to regain their self-confidence as professionals by acquiring Australian experience: "This is both teaching and learning because while doing this, I can socialise with my compatriots and Australians as well" (UMT3). The Ukrainian community school was a safe translingual space for the UMTs. As UMT3, who worked there as a volunteer teacher of English, described the school:

> It's [an] ideal place for all the Ukrainians, not only for me, just to feel this community, to find yourself in their environment […]. You hear people speaking the Ukrainian language. You share your problems, […] not just problems, but whatever you experience here in the Ukrainian language.

Despite all the difficulties the UMTs have encountered in Australia seeking professional realisation, four out of five stayed in the teaching field. As UMT4 said, it required time to gain Australian experience and to prove her worthiness, but once the person was in the "system", it would become easier to pursue one's career:

> I guess I'm more than satisfied because I didn't even apply for the jobs this time, so I guess once you get into the system, and you show that you can actually work. Current employment found

> me. They were just looking for the tutor, and my profile was there from the last year of teaching Ukrainian, and they just asked me if I'm interested. I said, "I'm interested," and that's it.

Only UMT2 chose to leave the teaching field and to start a new career. Though satisfied with her new job, she still had regrets about abandoning teaching:

> If I knew what I know now six years ago before coming to Australia, I would probably consider something connected with teaching because as far as I know, there are lots of positions in schools, and there are not so many professionals who can actually do that… With some additional courses to adjust your education received back in Ukraine. I know that there are Ukrainians who continue to teach [in Australia].

8. Conclusions

Applying the identity work concept of Lepisto et al. (2015), it is possible to say that UMTs in this study navigated barriers on their pathways of professional transition by retaining, subtracting and adding to their professional identities. Some of them manage to stay in the Australian pedagogical environment by pursuing a university course or acquiring a tertiary education certificate in Australia, thus adding to their professional identities developed overseas.

Others abandon their teaching professions, fully subtracting their identities and developing a new professional identity. Certain reasons for this new identity entail cases of accentism and translingual name discrimination reported by migrant teachers. Language-related prejudice often induces anxiety and hampers the cognitive capacities of those affected, as well as their interactions with individuals outside their social group (Birney et al., 2020). Ukrainian migrants experience the "glass ceiling effect" (Harrison, 2013, p. 200) because their accent and names act as an invisible barrier preventing them from finding jobs aligned with their training and experience. Since they have been accustomed to using Standard English since their university years, UMTs tend to feel comfortable communicating within academic circles. However, many have encountered barriers in everyday communication where colloquial Australian English is involved.

According to our findings, the concept of native speakerism still prevails in the ESL field, supported by both native and non-native speakers of English (Lai, 2021). NESTs speaking English as their first language are automatically considered the best teachers of English. Conversely, NNESTs have to prove their professional worthiness as they are considered novice teachers in the Australian pedagogical environment. This is reflected in unfair employment practices resulting in highly-qualified migrant teachers being unable to find jobs aligned with their expertise. Teacher registration requirements and IELTS testing, as well as other requirements, become a source of stress and anxiety, constituting a major problem for non-native English speakers in general (Aichhorn & Puck, 2017) and introducing uncertainty into UMTs' lives.

The aforementioned experiences can result in UMTs feeling self-doubt as English teachers in the Anglophone country, forcing some of them to leave the teaching field and pursue what they

deem a safer career in Australia. Others demonstrate their resilience and find employment in the teaching field, for example, by volunteering as teachers or teacher assistants. Thus, UMTs manage to continue teaching and overcome their lack of confidence as professionals. They retain their teaching identities by reinforcing their previous skills and using their vast pedagogical experience obtained overseas, even though their working positions in Australia are likely to be lower than those they had in their home country. Therefore, having an educational course in Australia is essential for any NNEST's successful career development.

Finally, a limitation of this study is that it embraces only UMTs. The research findings are also relevant for teachers from other ethnic groups, especially those from non-English-speaking countries, who encounter similar obstacles via their professional pathways in Australia; therefore, comparing the UMTs' lived experiences of professional transition in Australia with those of other migrant teachers could be a further research trajectory. Meanwhile, the findings of this study could potentially aid in crafting recommendations for pertinent institutions to enhance the migrant experience in the future.

References

Aichhorn, N. & Puck, J. (2017). "I just don't feel comfortable speaking English": Foreign language anxiety as a catalyst for spoken-language barriers in MNCs. *International Business Review*, *26*(4), 749–763.

Alieksieienko, T. (2020). Opportunities and risks of construction of a new identity in internally displaced persons (migrants). *Theoretical and Methodical Problems of Children and Youth Education*, *1*(24), 4–16.

Baratta, A. (2018). *Accent and teacher identity in Britain: Linguistic favouritism and imposed identities*. Bloomsbury Publishing.

Beesniova, C. (2017). "She's our spy": Power and positionality in studying Lithuanian teacher communities. In I. Silova, N.W. Sobe, A. Korzh & S. Kovalchuk (Eds.), *Reimagining utopias: bold visions in educational research* (pp. 15–31). Sense Publishers.

Birney, M. E., Rabinovich, A., Morton, T. A., Heath, H. & Ashcroft, S. (2020). When speaking English is not enough: The consequences of language-based stigma for non-native speakers. *Journal of Language and Social Psychology*, *39*(1), 67–86.

Brady, J. (2015). Dialect, power and politics: Standard English and adolescent identities. *Literacy*, *49*(3), 149–57.

Burke, C. T. (2014). *Biographical narrative interview method: Tracing graduates' futures*. SAGE Publications.

Clandinin, D. J. (2016). *Engaging in narrative inquiry*. Routledge.

Cohen, L., Manion, L. & Morrison, K. (2018). *Research methods in education*. Taylor & Francis Group.

Copland, F., Mann, S. & Garton, S. (2020). Native-English-speaking teachers: Disconnections between theory, research, and practice. *TESOL Quarterly*, *54*(2), 348–374.

Creswell, J. W. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

De Fina, A. (2015). Narrative and identities. In A. De Fina & A. Georgakopoulou (Eds.), *The handbook of narrative analysis* (pp. 352–368). Wiley.

Department of Home Affairs. (2022). *Ukraine visa support*. https://www.homeaffairs.gov.au/help-and-support/ukraine-visa-support#:~:text=Since%2023%20February%202022%20the,have%20since%20arrived%20in%20Australia

Dinesen, P. T., Schaeffer, M. & Sønderskov, K. M. (2020). Ethnic diversity and social trust: A narrative and meta-analytical review. *Annual Review of Political Science*, *23*, 441–465.

Dovchin, S. (2022). *Translingual discrimination*. Cambridge University Press.

Dryden, S. & Dovchin, S. (2021). Accentism: English LX users of migrant background in Australia. *Journal of Multilingual and Multicultural Development*, *5*, 1–13.

Ennerberg, E. & Economou, C. (2021). Migrant teachers and the negotiation of a (new) teaching identity. *European Journa*l *of Teacher Education*, *44*(4), 587–600.

Ershadi, F., Nazari, M. & Chegenie, M. S. (2024). Native speakerism as a source of agency – related critical incidents: Implications for non-native English teachers' professional identity construction. *System*, *120*, 1–12.

Fedyuk, O. & Kindler, M. (Eds.). (2016). *Ukrainian migration to the European Union: Lessons from migration studies*. Springer.

Gilmour, L., Klieve, H. & Li, M. (2018). Culturally and linguistically diverse school environments – exploring the unknown. *Australian Journal of Teacher Education*, *43*(2), 172–189.

Harrison, G. (2013). "Oh, you've got such a strong accent": Language identity intersecting with professional identity in the human services in Australia. *International Migration*, *51*(5), 192–204.

Holliday, A. (2005). *The struggle to teach English as an international language*. Oxford University Press.

Ibrahim, N., Hammed, H., Zaidan, A., Zaidan, B., Albahri, O. S., Alsalem, M. et al. (2019). Multi-criteria evaluation and benchmarking for young learners' English language mobile applications in terms of LSRW skills. *IEEE Access*, *7*, 146620–146651.

Jenkins, J. (2019). English medium instruction in higher education: The role of English as lingua franca. In X. Gao (Ed.), *Second handbook of English language teaching* (pp. 91–108). Springer.

Kang, S., DeCelles, K., Tilcsik, A. & Jun, S. (2016). Whitened résumés: Race and self-presentation in the labor market. *Administrative Science Quarterly*, *61*(3), 469–502.

Kartch, F. (2017). Narrative interviewing. In M. Allen (Ed.), *The SAGE encyclopedia of communication research methods* (pp. 1072–1075). SAGE.

Keaney, G. (2016). NEST schemes and their role in English language teaching: A management perspective. In F. Copland, S. Garton & S. Mann (Eds.), *LETs and NESTs: Voices, views and vignettes* (pp. 129–150). British Council.

Kiramba, L. K. & Harris, V. J. (2019). Navigating authoritative discourses in a multilingual classroom: Conversations with policy and practice. *TESOL Quarterly, 53*(2), 456–481.

Korstjens, I. & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and publishing. *European Journal of General Practice*, *24*(1), 120–124.

Laham, S. M., Koval, P. & Alter, A. L. (2012). The name-pronunciation effect: Why people like Mr. Smith more than Mr. Colquhoun. *Journal of Experimental Social Psychology*, *48*, 752–756.

Lai, M. L. (2021). English linguistic neo-imperialism – a case of Hong Kong. *Journal of Multilingual and Multicultural Development*, *42*(4), 398–412.

Legislation of Ukraine. (2014). *Law of Ukraine on Higher Education*. https://zakon.rada.gov.ua/laws/show/1556-18/paran1165?lang=en#n1165

Lepisto, D. A., Crosina, E. & Pratt, M. G. (2015). Identity work within and beyond the professions: Toward a theoretical integration and extension. In A. Desilva & M. Aparicio (Eds.), *International handbook of professional identities* (pp.11–37). Scientific & Academic Publishing.

Markus, A. (2014). Attitudes to immigration and cultural diversity in Australia. *Journal of Sociology*, *50*(1), 10–22.

Marom, L. (2019). From experienced teachers to newcomers to the profession: The capital conversion of internationally educated teachers in Canada. *Teaching and Teacher Education*, *78*, 85–96.

Maydell, E. (2020). "And in Israel we became Russians straight away": Narrative analysis of Russian-Jewish identity in the case study of double migration. *Narrative Inquiry, 30*(2), 404–426.

Mertova, P. & Webster, L. (2019). *Using narrative inquiry as a research method: An introduction to critical event narrative analysis in research, teaching and professional practice*. Routledge.

Oleinikova, O. (2020). *Life strategies of migrants from crisis regimes: Achiever or survivor?* Springer.

Oliver, R., Rochecouste, J. & Nguyen, B. (2017). ESL in Australia – A chequered history. *TESOL in Context 26*(1), 7–26.

Papathomas, A. (2017). Narrative inquiry: From cardinal to marginal… and back? In B. Smith & A. C. Sparkes (Eds.), *Routledge handbook of qualitative research in sport and exercise* (pp. 37–48). Routledge.

Park, E. S. & Lee, H. (2022). "I want to keep my North Korean accent": Agency and identity in a North Korean defector's transnational experience of learning English. *TESOL quarterly, 56*(1), 19–40.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and practice* (4th ed.). SAGE Publications.

Phillipson, R. (2012, March 13). Linguistic imperialism alive and kicking. *The Guardian*. https://www.theguardian.com/education/2012/mar/13/linguistic-imperialism-english-language-teaching

Radhouane, M., Akkari, A. & Consuelo, G. M. (2022). Understanding social justice commitment and pedagogical advantage of teachers with a migrant background in Switzerland: A qualitative study. *Journal for Multicultural Education, 16*(2), 159–170.

Rangan, P., Saxena, A., Srinivasan, R. T. & Sundar, P. (Eds.). (2023). *Thinking with an accent: Toward a new object, method, and practice*. University of California Press.

Rivers, D. J. (2018). Speakerhood as segregation: The construction and consequence of divisive discourse in TESOL. In B. Yazan & N. Rudolph (Eds.), *Criticality, teacher identity, and (in)equity in English language teaching* (pp. 179–197). Springer.

Schwandt, T. A., Lincoln, Y. S. & Guba E. G. (2007). Judging interpretations: But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. *New Directions for Evaluation*, *114*, 11–25.

Schwartz, S. J., Luyckx, K. & Vignoles, V. L. (Eds.). (2011). *Handbook of identity theory and research*. New Springer.

Sterne, J. (2022). *Diminished faculties: A political phenomenology of impairment*. Duke University Press.

Stets, J. E. & Serpe, R. T. (2013). Identity theory. In J. DeLamater & A. Ward (Eds.), *Handbook of social psychology* (pp. 31–60). Springer.

Tankosić, A. (2022). *Linguistic diversity and disparity in the periphery*. [Unpublished Doctoral dissertation]. Curtin University, Western Australia. https://espace.curtin.edu.au/handle/20.500.11937/91773

Teacher Registration Board of Western Australia. (2024). *Overseas qualified teachers*. https://www.trb.wa.gov.au/Teacher-Registration/Becoming-registered/Overseas-qualified

Trudgill, P. (2002). *Sociolinguistic variation and change*. Edinburgh University Press.

Tupas, R. (2021). Fostering translingual dispositions against unequal Englishes. *English in Education*, *55*(3), 222–238.

UNESCO (2023). International Mother Language Day: UNESCO calls on countries to implement mother language-based education. https://www.unesco.org/en/articles/international-mother-language-day-unesco-calls-countries-implement-mother-language-based-education

Wang, S., Cai, W., Sun, Q., Martin, C., Tewari, S., Hurley, J., Amati, M., Duckham, M. & Choyet, S. (2023). Landscape of multiculturalism in Australia: Tracking ethnic diversity and its relation with neighbourhood features in 2001–2021. *Applied Geography*, *160*, 1–12.

Wright, N. (2022). (Re)production of symbolic boundaries between native and nonnative teachers in the TESOL profession. *Asian-Pacific Journal of Second and Foreign Language Education*, *7*(1), 1–13.

Zeng, J., Ponce, A. R. & Li, Y. (2023). English linguistic neo-imperialism in the era of globalization: A conceptual viewpoint. *Frontiers in Psychology*, *14*, 1–9.

Zhao, X. & Biernat, M. (2017). “Welcome to the U.S.” but “change your name”? Adopting Anglo names and discrimination. *Journal of Experimental Social Psychology*, *70*, 59–68.

***

Larysa Chybis is an interpreter, and a PhD student at Curtin University, Western Australia. Her research interests include second and foreign language teaching, migrant cultural studies, comparative linguistics, English literature and environmental studies. Also, Ms Chybis has been

dedicating her time to teaching English to displaced Ukrainians and other migrants, helping them to settle down in Australia. Ms Chybis obtained her master's degree in Ukraine, where her major professional activity was teaching English in tertiary educational institutions.

Dr **Sally Lamping** has spent over twenty years as a teacher, teacher educator, and researcher in primary, secondary, and adult education contexts across the world. A large portion of her research is on the secondary English classroom and youth identities, with a specific focus on migrant youth and how schools can become enabling spaces for students and teachers. She was a 2015 U.S. Fulbright Core Senior Research Scholar in Adelaide, South Australia, where she worked and conducted research alongside newly arrived students in Australia's only stand-alone New Arrival Program for adolescents. She is currently the project lead on three multifaceted Critical Participatory Action Research projects with migrant communities in Local Government Areas of Perth; the projects focus on how learning happens in situational community-driven contexts that can inform sustainable local government initiatives.

Prof. **Toni Dobinson** is the Coordinator of the Post Graduate Programs in Applied Linguistics including the MA Applied Linguistics, the MTESOL and the Graduate Certificate in TESOL. She is also the Discipline lead for Applied Linguistics/TESOL and Languages. She has many years of teaching and research experience overseas in Egypt and the Sultanate of Oman. She has coordinated and taught the MA Applied Linguistics for over 20 years including offshore in Ho Chi Min City, Vietnam. She also teaches in the Language and Diversity unit of the BEd Early Childhood/Primary and supervises Higher Degree by Research students. She has published in the areas of language education, classroom research, translanguaging and linguistic racism. The title of her PhD was *Investigating the teaching and learning experiences of Asian postgraduate students and their lecturers in Australia and Vietnam*.

Dr **Kathryne Ford** is Deputy Director of Graduate Research in Curtin University's School of Education. She previously managed the Australian Literary Studies academic journal, and, in addition to her Curtin role, Kathryne is currently a researcher in the School of Literature, Languages and Linguistics at the Australian National University, where she completed her PhD. Kathryne has a wide array of research and teaching interests, including biofiction, literacy, academic skills, creative writing, life-writing, cultural memory, gender studies, art, and narratology. Her articles on these topics have appeared in a/b: *Auto/Biography Studies*, *The Dickens Quarterly*, *The Wilkie Collins Journal*, and *The Australasian Journal of Victorian Studies*. Kathryne's current research projects investigate neo-historical biofiction and memory studies.

JASMINA JELČIĆ ČOLAKOVAC[1] DOI: 10.15290/CR.2024.45.2.04
University of Rijeka, Faculty of Maritime Studies, Croatia
https://orcid.org/0000-0002-1241-1283
IRENA BOGUNOVIĆ
University of Rijeka, Faculty of Maritime Studies, Croatia
https://orcid.org/0000-0002-2956-7014

Putting languages into perspective: A comprehensive database of English words and their Croatian equivalents

Abstract. Numerous studies have addressed the issue of English words in the context of their adaptation, but there still exists the need for a systematic perspective on English words in terms of their number and frequency of appearance. This article will outline the procedure behind the compilation process of unadapted English words in the Croatian language with a comprehensive description of the final product – an open-access database of single- (SWE) and multi-word (MWE) English expressions extracted from Croatian web corpora (*ENGRI* and *hrWaC*) by means of computational-linguistic tools and manual extraction. The final version of the database contains 2,982 English words in their unadapted form (e.g. *blockbuster*), and 18 words which appear with English orthographic properties in combination with Croatian inflectional affixes (e.g. *downloadati*). Each SWE and MWE entry in the database is accompanied with frequencies of appearance in both corpora as well as its Croatian equivalent where available (29.58% of all entries are listed without an equivalent). The database serves as the first systematic representation of English words in Croatian and provides an indispensable tool for further research into the phenomenon while at the same time opening the door to a new line of research – cognitive processing of English words in Croatian.

Keywords: English words in Croatian, language borrowing, corpus search, database compilation, anglicisms

1. Introduction

Borrowing from English has been documented in many languages. Words and expressions borrowed from English have been investigated in Spanish (e.g. Alvarez-Mellado, 2020), Italian

1 Address for correspondence: Faculty of Maritime Studies, Foreign Languages Department, Studentska 2, 51 000 Rijeka, Croatia. E-mail: jasmina.jelcic@pfri.uniri.hr

(e.g. Pulcini et al., 2012), Norwegian (e.g. Greenall, 2005), Slovenian (e.g. Čepon, 2017), Czech and Slovak (e.g. Entlová & Mala, 2020), Japanese (e.g. Kay, 1995) and South Korean (e.g. Rüdiger, 2018), to mention some of them. Croatian has also become highly receptive to borrowing from English (Mihaljević Djigunović & Geld, 2003). As a result, many English loanwords have become part of Croatian everyday communication (e.g. Nikolić-Hoyt, 2005). The prestigious status of a donor language (e.g. Crystal, 2003) reduces the tendencies of borrowed words to fully adapt to the rules of the recipient language (e.g. McKenzie, 2010; Nikolić-Hoyt, 2005).

Borrowed words are generally described in terms of the degree of their adaptation to the recipient language (e.g. Görlach, 2002; Entlová & Mala, 2020) or their inclusion in the language (e.g. Kay, 1995; Međeral, 2016). A distinction is made between words which have adapted, fully or partially, to the recipient language and those which occur in an original, unadapted form (e.g. *event*, *freelancer*, *bodybuilder*, etc.). Terminology related to unadapted English loanwords is not unified, so terms like 'raw anglicisms' (e.g. Kavgić, 2013), 'English loanwords' (e.g. Görlach, 2002; Kay, 1995; Rüdiger, 2018), 'foreign words' (e.g. Međeral, 2016; Muhvić-Dimanovski & Skelin Horvat, 2006) and 'pseudoanglicisms' (e.g. Filipović, 1990) can be found.

This paper focuses on the latter category, i.e. words borrowed from English which retain the original properties of the donor language, and sometimes take Croatian affixes (e.g. *eventi* (m. nom. pl.), *freelancerima* (m. dat. pl.), *bodybuildera* (m. gen. sg.), etc.). Such words have not become an integral part of Croatian and are perceived as foreign by native speakers, so the term 'foreign words' seems appropriate. For the purpose of precision, the term 'English words' will be used (e.g. Brdar, 2010; Ćoso & Bogunović, 2017).

Borrowed words have long been a subject of discussion among Croatian linguists (Muhvić-Dimanovski & Skelin Horvat, 2006), who generally recommend the use of native words (e.g. Hudeček & Mihaljević, 2005). There are several ways to deal with borrowed words: using multi-word expressions and descriptions, using an existing word and giving it a new meaning, or introducing new words and calques. However, it seems that not all such solutions have been accepted among Croatian speakers (e.g. Drljača, 2006; Patekar, 2019), especially in domains like show business and information technology (e.g. Drljača Margić, 2014). Multi-word expressions and descriptions are often complex to use (e.g. Drljača, 2006). For example, according to the website *Bolje je hrvatski!* (*bolje.hr*), the English word *software* is translated as *programska podrška* (Eng. 'program support'), and *developer* as *razvojni inženjer* (Eng. 'development engineer'). The complexity of these solutions is best illustrated by the translation of the syntagm *software developer* as *razvojni inženjer programske podrške* (Eng. 'program support development engineer'). Giving a new meaning to an already existing word can result in insufficient precision (Drljača, 2006), as in *spravica* (Eng. 'small device') for *gadget* (*bolje.hr*). Finally, the process of introducing a new word or calque is usually slow (e.g. Muhvić-Dimanovski & Skelin Horvat, 2008). For example, the English word *selfie* gained worldwide popularity in 2012, while the Croatian equivalent *sebić* was proposed in 2014 (Halonja & Hudeček, 2014).

In Croatia, a vast body of research has investigated the phenomenon of English words using different theoretical approaches and methods (e.g. Ćoso & Bogunović, 2017; Drljača Margić, 2014; Filipović, 1990; Patekar, 2019). However, most researchers either focus on selectively chosen English words (e.g. Ćoso & Bogunović, 2017; Patekar, 2019) or rely on small-scale, domain-specific corpora (e.g. Brdar, 2010; Hudeček & Mihaljević, 2005). What seems to be neglected is a data-driven approach. One possible reason for that could be the fact that the development of Croatian computational linguistic tools and resources lagged behind those of other languages in the past (e.g. Tadić et al., 2012). However, this is now changing and some new language technologies have been developed in the last decade (Tadić, 2022). Aside from traditional dictionaries (e.g. Filipović, 1990; Görlach, 2002), new resources have emerged. For example, the above-mentioned website *Bolje je hrvatski*!, developed by the *Institute for Croatian language and linguistics*, selectively records the intake of foreign words into Croatian and proposes native equivalents. Borrowed words, including some English words, can also be found in an online dictionary of neologisms (Muhvić-Dimanovski et al., 2016). On the other hand, *Kontekst.io* searches the Croatian web corpus, *hrWaC* (Ljubešić & Klubička, 2016) to find a specific word. The results include information about the word's frequency as well as the frequencies of similar words. Word frequency can also be obtained by searching for a specific word in the available corpora via the platform *Sketch Engine* (Kilgarriff et al., 2004). The results are presented in context, and various options are available to filter them out (e.g. English words occurring in English contexts, names, etc.). However, this method cannot be used to create lists of English words, as the existing corpora are linguistically processed (e.g. tokenized, lemmatized, morphosyntactically tagged, etc.) according to the rules of the Croatian language.

In other languages, researchers have used different methods for the extraction of anglicisms and English words from corpora. Some authors opted for manual search (e.g. Luján García, 2017), and others used the available tools and resources or created new ones (e.g. Alex, 2005; Andersen, 2012). For example, an unsupervised system, based on the idea that there is a relation between Google search results and language membership, was developed for the classification of anglicisms in German (Alex, 2005). Another approach combined lexicon lookup with character N-grams (e.g. Furiassi & Hofland, 2007). Supervised machine learning methods in combination with N-grams has also yielded reliable results (e.g. Alvarez-Mellado, 2020; Serigos, 2017), and it was used to create the Database of English words in Croatian (Bogunović & Kučić, 2022).

The *Database of English words in Croatian* (Bogunović & Kučić, 2022) contains 9,453 English words, some of which (e.g. *summit*, *vintage*, *benefit*) originate from other languages. Although some authors (e.g. Filipović, 1990) state that even words that are not English in origin but were borrowed from English can be considered English loanwords, establishing each word's etymology was not the goal of Bogunović and Kučić's work. The Database represents the result of algorithmic classification and manual evaluation of word lists produced by the algorithm. The results are publicly available on *Figshare.com* as an open source of data. However, it does not provide any information about the availability of Croatian translational equivalents and their frequencies.

Moreover, the database only lists extracted English words with their frequencies, without further elaboration of context-related problems such as polysemy, interlingual cognates, proper names, etc.

The following research aims to fill these gaps by further elaborating the Database of English words in Croatian. The paper presents the results of such an endeavor.

2. Method

The Database of English words and their Croatian equivalents (Bogunović, Jelčić *Čolakovac* & Borucinsky, 2022, hereinafter: the Database) presents an elaboration of Bogunović and Kučić's (2022) database, based on the *ENGRI* corpus (Bogunović et al., 2021; Bogunović & Kučić, 2021), which contains texts from the 12 most popular Croatian news portals between 2014 and 2020. The Database was further updated with data from both *ENGRI* and *hrWaC 2.2* (Ljubešić & Klubička, 2016), built by crawling the .hr top-level domain in 2011 and again in 2014, using the SketchEngine (SkE) platform (Kilgarriff et al., 2004).

2.1. Manual search and evaluation

Manual evaluation of corpus data was used to eliminate the entries from Bogunović and Kučić'l s (2022) database which had appeared in the corpora either in embedded English texts or as part of an English multi-word expression (MWE), the phrase constituents of which the algorithm recognized as single-word (SWEs) entries. The issue of English words appearing in English contexts rather than Croatian sentences was ultimately resolved through SkE search and Xf tagger filtration.

Manual search and evaluation was, however, indispensable in resolving the MWE issue, along with a number of problems which were brought to our attention during corpus search (Jelčić Čolakovac & Borucinsky, 2023). These issues include:

1. the disambiguation of proper names and common nouns (e.g. *PlayStation* as a company vs. *playstation* as a term for a gamer console, etc.);
2. the absence of diacritics from Croatian words appearing in web-crawled sources (e.g. Cro. *vaše* (pro., 2nd pers. pl.) 'yours' vs. Eng. *vase* 'decorative container', etc.);
3. meaning disambiguation (e.g. Cro. *gem* (m. nom. sg.) 'term in a tennis scoring match' vs. Eng. *gem* 'a precious stone', etc.);
4. inflection of Croatian word classes (e.g. Cro. *elaborate* (m. acc. pl.) 'a written elaboration' vs. Eng. *elaborate* (v.) 'to explain' or Eng. *elaborate* (adj.) 'planned in detail', etc.);
5. false cognates (e.g. Cro. *file* (m. nom. sg.) 'chicken breast' vs. Eng. *file* 'document', etc.)
6. adapted English forms (e.g. Cro. *bend* (n.) 'band, a group of musicians' vs. Eng. 'bend' (v.) 'to turn or force from straight or even to curved or angular', etc.).

Once the list of English words was filtered using corpus tools and manual search, our effort was directed towards providing Croatian translational equivalents for each entry in the Database.[2]

2 The term 'entry' will be used interchangeably throughout the paper to refer to both single-word (SWE) and multi-word (MWE) expressions from the Database.

Published sources (dictionaries, books, articles, etc.) on the topic of English loanwords in Croatian served as the stepping-stone to finding adequate equivalents, and, if these proved insufficient, web sources and both corpora were used to aid the search.

Table 1 lists examples from the Database and sources which were used to identify their Croatian equivalents.

Table 1. An exemplification of sources for Croatian equivalents

Entry	Croatian equivalent	Source
ability	*sposobnost* (f. nom. sg.)	Bujas (2019)
afterparty	*zabava nakon posla* (f. nom. sg.; prep.; m. dat. sg.)	Bolje (`https://bolje.hr/`)
bookmark	*straničnik* (m. nom. sg.)	Muhvić-Dimanovski and Skelin Horvat (2008)
	knjižna oznaka (f. adj. sg.; f. nom. sg.)	Glosbe (`https://hr.glosbe.com/`)
	dočitnica (f. nom. sg.)	Wiktionary (`https://www.wiktionary.org/`)
composite	*presjek* (m. nom. sg.)	hrWac
	kompozitan (m. adj. sg.)	hrWac
dongle	*hardverski ključ* (m. adj. sg.; m. nom. sg.)	Glosbe (`https://hr.glosbe.com/`)
grooming	*uređivanje pasa* (n. nom. sg.; m. gen. pl.)	Bolje (`https://bolje.hr/`)
homepage	*početna stranica* (f. adj. sg.; f. nom. sg.)	Bujas (2019)
	službena stranica (f. adj. sg.; f. nom. sg.)	Glosbe (`https://hr.glosbe.com/`)
blind tasting	*kušanje na slijepo* (n. nom. sg.; prep.; n. adv. sg.)	hrWac
acting coach	*učitelj glume* (m. nom. sg.; f. gen. sg.)	Bujas (2019)
	učiteljica glume (f. nom. sg.; f. gen. sg.)	
corkage fee	*naknada za služenje* (f. nom. sg.; prep.; n. nom. sg.)	ENGRI
	čeparina (f. nom. sg.)	Glosbe (`https://hr.glosbe.com/`)
remote reality	*udaljena stvarnost (f. nom. sg.)*	ENGRI

2.2. Semantic analysis of sentential context

To evaluate the context in which a particular English word is used in Croatian, the two corpora had to be swept for representative samples of sentential context using the SkE search tools. For those entries appearing in multiple contexts, the predominant context was used in assigning the word to a specific area of human activity, i.e. the semantic field it is usually associated with in the Croatian corpora. The word *design* (*RF* = 7.1402) is one such example, which also appears in contexts related to information and communications technology (ICT), but is predominantly used in contexts related to the fashion industry. Other examples include English words such as *abuse* (*RF* = 0.2683; appearing in law and politics-related context, but predominantly in ICT-related context) and *combat* (*RF* = 0.5191; appearing in sentential contexts related to war, but predominantly in sport and gaming contexts). For those words appearing across multiple contexts with similar frequencies, or whose semantic field could not be determined due to the word's generic reference, the 'OTHER' category was introduced. Such instances include SWEs like *review* (*RF* = 0.7497), *progress* (*RF* = 0.5623), *position* (*RF* = 1.5416) and *reach* (*RF* = 0.4572), and MWEs like *against type* (*RF* = 0.0095), *boom effect* (*RF*= 0.0019) and *extreme ways* (*RF*= 0.0087), which appear in the corpora in various contexts. Based on in-depth semantic analysis of sentential context in which the English words appeared in the Croatian corpora, 12 semantic categories and 12 subcategories have been introduced (Table 2).

Table 2. Representation of semantic categories (n = 12) and subcategories (n =12) for English words in the Database

	Category	Description	Examples
(1)	**ANT**	relating to animals, plants, and non-human entities with human traits	*beast, spider, queen bee*
(2)	**ART**	relating to entertainment and show business, and branches of human creative activities, such as music, dance, literature, etc.	*classic, comeback, casting director*
Subcategory			
	MUSIC	relating to the music industry	*airplay, orchestra, jam session*
	TV	relating to tv, news, and film industry	*binge, spoiler, body horror*
(3)	**PEOPLE**	relating to people, human behavior and activity, and social phenomena in general	*gay, teenager, attention whoring*
Subcategory			
	LANG	relating to language and linguistic phenomena, metaphor, and idiomatic language	*actually, anyway, be on fire*

	Category	Description	Examples
	WAR	relating to combat and human conflict in general	*battle, raid, ground zero*
	LP	relating to government, law, and politics	*e-government, council, cell block*
(4)	**BUSINESS**	relating to business and economy, finance, money, and the world of work in general	*brownfield, offset, debt equity swap*
Subcategory			
	COMMERCE	relating to the act of buying and/or selling, product advertising, and consumerism in general	*delivery, tester, customer loyalty*
(5)	**TECH**	relating to technology and operation of machinery	*clutch, joystick, driver screen*
Subcategory			
	ICT	relating to information and communications technology, Internet, and computer science	*feed, inbox, big data*
	TRANSPORT	relating to means of transport and transport-connected activities	*cargo, landing, economy class*
(6)	**SCIENCE**	relating to science and scientific activity	*molecular, nuclear, case study*
Subcategory			
	EDUCATION	relating to educational activities	*academy, e-learning, action learning*
(7)	**FASHION**	relating to clothing, make-up, style, and the beauty business	*casual, styling, dress code*
(8)	**FOOD**	relating to food and drink, and the act of dining and diet in general	*beef, drive-in, blind tasting*
(9)	**HEALTH**	relating to health, medicine, and the human body	*operation, pill, blood aging*
Subcategory			
	SPORT	relating to sport and games	*draft, playmaker, alpine skiing*
(10)	**TOURISM**	relating to the tourist business and travel for pleasure	*all-inclusive, booking, foot holiday*
Subcategory			
	NATURE	relating to environment and ecology	*emission, winter, hot spring*

	Category	Description	Examples
	LOC	relating to specific places and localities	*penthouse, room, food corner*
(11)	**QUANTITY**	relating to quantity, size, position or duration	*low-level, zero, long term*
(12)	**OTHER**	words with generic references and/or words appearing across multiple contexts	*ancient, progress, free choice*

The proposed categorization is based on the Croatian contexts in which the English words appear, and can by no means be taken to reflect the semantic contexts in which these words are regularly used in English. We would also like to stress that only the most representative semantic categories have been identified. Furthermore, subcategories have been assigned based on available corpus evidence and where repetitive overlap between semantic categories has been observed (e.g. NATURE and LOC have been categorized under TOURISM since a considerable number of words belonging to the two subcategories have repeatedly appeared in contexts relating to tourism and travel, albeit with lower frequencies than in their assigned subcategories).

Finally, after resolving problems through manual search and human evaluation, finding translational equivalents in Croatian, and assigning semantic categories to each entry, the Database has been published as an open-source linguistic resource, with the representation of data in tabular form (row per entry) (Figure 1).

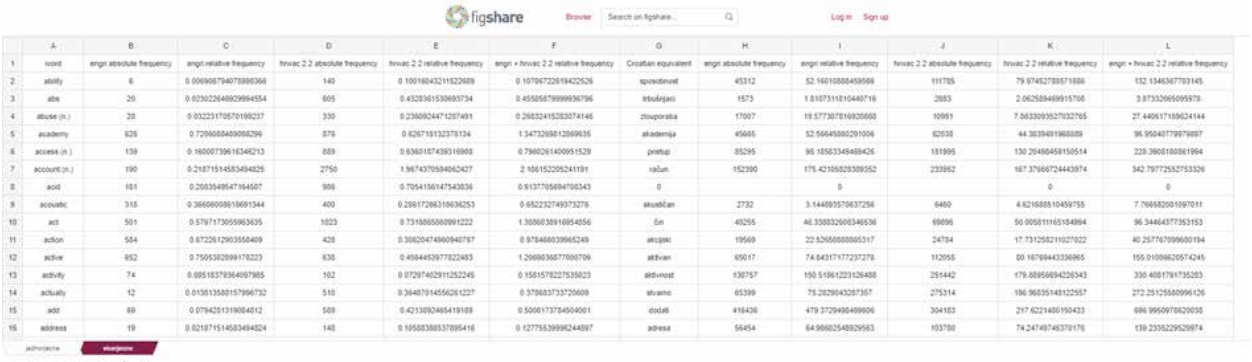

Figure 1. The Database available as open source on *Figshare.com*

3. Results and discussion

The Database contains 2,964 English words and expressions which appear in Croatian texts in their original, unadapted form (e.g. *blockbuster*, *cyberbullying, shopping, zombie, skin*, etc.) and 18 words with English orthographic properties in combination with Croatian inflectional affixes (e.g. *downloadati* (v.t., inf.) ‘to download’, *managerica* (f. nom. sg.) ‘female manager’, etc.).

3.1. Word frequencies

Each database entry is accompanied with a Croatian equivalent if the latter exists in the Croatian language. Absolute frequencies expressing the total number of corpus occurrences for each entry in the database and relative frequencies expressing the proportion of each entry's occurrence in the entire corpus (absolute frequency divided by the total number of words per corpus) are listed for both the English expression and, if applicable, its equivalent. *ENGRI* and *hrWaC 2.2* corpora served as the starting point for the calculation of frequencies which are represented in the Database both per corpus and combined: *ENGRI* absolute frequency (*E*af), *ENGRI* relative frequency (*E*rf), *hrWaC* absolute frequency (*H*af), and *hrWaC* relative frequency (*H*rf). The Database also provides data on combined relative frequencies (*RF*) for both corpora.

Only five entries have been shown to appear in the corpora more than 100,000 times (*web*, *real*, *blog*, *show*, and *post*), while 85 words (2.85%) appear more than 10,000, and less than 100,000 times. SWEs belonging to this frequency band include *link*, *fan*, *e-mail*, *online*, *net*, *mail*, *rock*, *jazz*, etc., with only one MWE appearing in the corpora more than 10,000 times (*big brother*) (cf. Table 2). In total, 709 SWEs and 27 MWEs appear between 1,000 and 10,000 times, which accounts for 24.68% of the Database. If we take the bottom-up perspective on frequencies, 41.78% of all Database entries are recorded 100 times or less in the corpora (184 SWEs and 1062 MWEs respectively), with some MWE entries (e.g. *age verification*, *all girl band*, *anti age effect*, *anti stain effect*, *appearance fee* (*RF* = 0.0012), etc.) and only five SWE entries (*mapmatching*, *mastershot*, *spraypainting* (*RF*= 0.0012), and *personalization* (*RF* = 0.0007)) appearing only once in the Croatian context.[3]

Table 3 illustrates the 10 entries with the highest combined relative frequencies (*RF*) in the Database.

Table 3. Database entries with the highest relative frequencies on the SWE and MWE lists

SWEs					
Entry	**E*af***	**E*rf***	**H*af***	**H*rf***	***RF***
real	86346	99.3957	46730	33.4321	132.8278
web	27648	31.8265	116672	83.4708	115.2973
show	69705	80.2397	36043	25.7863	106.0260
blog	12710	14.6309	112350	80.3787	95.0096
post	18565	21.3708	85431	61.1200	82.4908

3 We would like to note here that all Database entries reflect the spelling of the word(s) as it was used in the Croatian context, which does not necessarily adhere to the standards of the spelling rules for the English language (e.g. *anti age effect* instead of *anti-age effect*, etc.). The same approach was followed in sorting the entries into SWEs and MWEs (e.g. *mapmatching* rather than *map matching*, etc.).

SWEs					
fan	40273	46.3596	39346	28.1494	74.5089
link	6037	6.9494	79778	57.0757	64.0251
online	25053	28.8393	43498	31.1198	59.9592
e-mail	19318	22.2376	49698	35.5555	57.7931
mail	15473	17.8115	42794	30.6162	48.4277
MWEs					
big brother	9657	11.1165	4833	3.4577	14.5742
stand(-)up	2161	2.4876	2055	1.4702	3.9578
fast food	1755	2.0202	2532	1.8115	3.8317
triple(-)double	2920	3.3613	398	0.2847	3.6460
fair play	1606	1.8487	2036	1.4566	3.3053
single	815	0.9382	2329	1.6662	2.6044
made in	663	0.7632	2510	1.7957	2.5589
red carpet	285	0.3281	3061	2.1899	2.5180
open source	141	0.1623	2949	2.1098	2.2721
must have	828	0.9531	1596	1.1418	2.0950

3.2. Single-word and multi-word expressions

The categorization of English words into 1,728 single-word (Cro. *jednorječne*) (SWEs) and 1,254 multi-word (Cro. *višerječne*) expressions (MWEs) represents one of the two major elaborations of Bogunović and Kučić's (2022) database. The restrictions of the original algorithm (Bogunović & Kučić, *under review*), which produced word lists for both databases, prevented it from recognizing English MWEs in the web-crawled sources, hence turning manual evaluation and corpus search into necessary methodological steps in the compilation of our Database.

On the one hand, a detailed manual search of both *hrWaC* and *ENGRI* corpora revealed that many of the words which were initially tagged by the algorithm as SWEs were, in fact, part of an English MWE used in a Croatian context (such examples include English words like *flower* (appearing only as a constituent in the MWEs *flower power* and *flower fashion*) or *cat* (appearing only in MWEs *cat and mouse*, *cat person*, and *cat people*). On the other hand, further examination of corpus examples indicated that some English words were used in Croatian as either a SWE or part of an MWE. These words include, for example, *age* (appearing also in MWEs *age verification*, *anti-age (effect)*, and *coming of age*), *horror* (also in *body horror* and *shock horror*), or *zero*

(also in *ground zero*, *patient zero*, *size zero models*, *zero companies*, *zero hour contract*, and *zero waste*). These entries used as both SWEs and part of English MWEs in Croatian needed to be taken into consideration when absolute and relative frequencies were concerned; it was upon the evaluators to rely on KWIC (*key word in context*) searches in order to distinguish between the SWE and MWE frequencies for words appearing in both wordlists (e.g. the occurrences of the word *coffee* in the MWEs *coffee culture* and *ice coffee* needed to be subtracted from the overall frequencies for the SWE *coffee*). Once these issues had been resolved, the absolute and relative frequencies could be added to the Database for both SWEs and MWEs.

Compounds presented a particular challenge in the process of compilation since some items appeared in the Croatian texts in both hyphenated and non-hyphenated forms. If the English expression appeared in the corpora either as a single word or a hyphenated compound (e.g. *blu(e)-ray*, *all-in-one*, *talk-show*, *co-creation*, *all-inclusive*, *mid-range*, *one-on-one*, *speech-to-text*, *co-production*, *follow-up*, *drive-in*, *pet-friendly*, *ready-made*, etc.), it was categorized as a SWE. The MWE entries used in Croatian as hyphenated MWEs (e.g. *make(-)up artist*, *e(-)book reader*, *regional stand(-)up*, *pop(-)up corner*, etc.) were categorized under MWEs with the hyphen placed in parentheses in order to indicate its optionality. Finally, six entries were listed under both SWEs and MWEs since they appeared in the corpora with and without a hyphen, i.e. as a MWE (*triple(-)double*, *hi(-)tech*, *jet(-)ski*, *head(-)up*, *cut(-)out*, and *stand(-)up*). The final categorization yielded 62 hyphenated compounds on the SWE list, which constitutes 3.59% of the total number of single-word entries in the Database whereas the MWE list included 13 hyphenated entries (1.04% of the total number of multi-word entries). The SWE compound which most frequently appeared in the Croatian corpora is *e-mail*, with a combined relative frequency of 57.79, followed by *start-up* (*RF* = 11.43), *triple(-)double* (*RF* = 3.89), and *blu(e)-ray* (*RF* = 3.07).

3.3. English words with Croatian affixes

Apart from the inclusion of unadapted English words, the Database also lists English words which have taken on Croatian inflectional forms (0.60% of the total number of database entries, 18 entries in total), the majority of which are single-word entries (two inflected MWE entries have been recorded, namely *location manager(ica)* (*managerica*, f. nom. sg.) 'female location manager' and *teen seks comedy* (*seks*, m. nom. sg.) 'teen sex comedy'.

The largest portion of inflected words have taken on the Croatian inflectional suffix *-ica,* which denotes the female gender in Croatian and indicates the female doer of an activity (examples include: Cro. *sprinterica* (f. nom. sg.) 'a female sprinter'; Cro. *managerica* (f. nom. sg.) 'a female manager'; Cro. *youtuberica* (f. nom. sg.) 'a female youtuber'; Cro. *swingerica* (f. nom. sg.) 'a female swinger', etc.). The inflectional suffix was also recorded with *wagsica* (f. nom. sg.), even though the word in English may only refer to women (WAG is literally the acronym of 'wife and girlfriend' and, according to the *Cambridge Diction*ary, stands to denote 'a wife or girlfriend, especially of a well-known sports player'). The *-ica* suffix also appeared in *hoodica* (f. nom. sg. 'a hoodie'), where it does not refer to a female doer, but rather the female gender

of the noun in question, whereby feminine noun properties were added to the English word *hoodie*, probably due to meaning similarities with another Croatian word, *majica* (f. nom. sg., 'any type of T-shirt, blouse, or shirt').

Other Croatian inflectional suffixes which appeared with English words in the corpora include the nominal suffix *-anje* (*swinganje* (n. nom. sg.) 'the act of swinging'), the Croatian verbal suffix *-(a)ti* (*downloadati* (v.t., inf.) 'to download'; *googlati* (v.t., inf.) 'to google') and the adjectival/adverbial suffix *-no* (*maximalno* (n. nom. sg.) 'in the largest or greatest manner'). If an English word was used in the Croatian context in both its unadapted and inflectional form, the two words were listed as separate entries (such was the case with *download* and *downloadati*).

3.4. Croatian equivalents

The second major elaboration of Bogunović and Kučić's (2022) database lies in the addition of Croatian equivalents and their absolute and relative frequencies in both corpora. A total of 29.58% of all the entries in the Database are listed without a Croatian equivalent (296 SWEs and 586 MWEs), while 54 SWE entries and 28 MWE entries are listed with more than one possible equivalent. More than two Croatian equivalents are listed for 11 SWE entries (*bookmark*, *manager*, *managerica*, *maker*, *kickboxer*, *investor*, *hero*, *hater*, *stylist*, *rookie*, and *policy-maker*) and 3 MWE entries (*cloud computing*, *comedy club*, and *cooking class*).

Translational equivalents were found in Croatian for most entries in the Database (e.g. Eng. *ability/*Cro. *sposobnost*, Eng. *air guitar*/Cro. *zračna gitara*, Eng. *zombie*/Cro. *zombi*, Eng. *wild*/Cro. *divlji*, Eng. *winner*/Cro. *pobjednik*, Eng. *city*/Cro. *grad*). In those instances where the English word appeared in the Croatian context bearing more than one meaning, Croatian equivalents were listed separately to account for each of the word's meanings (e.g. Eng. *company*/ Cro. *kompanija* 'an organization that sells goods or services in order to make money'*, društvo* 'the fact of being with a person or people, or the person or people you are with'). Croatian equivalents for other meanings of *company* which are found in English are not listed in the Database since the word does not appear to be used in those senses (e.g. Eng. *company* 'a group of actors, singers, or dancers who perform together', 'a large group of soldiers', 'an organized group of young women who are guides', etc.). Similarly, a word was provided with the Croatian equivalent which would belong to the word category in which it was used in the Croatian context. This is to say, English words such as *update*/Cro. *posuvremeniti* (v.t., inf.), *edit*/Cro. *urediti* (v.t., inf.), *ski*/Cro. *skijati* (v.int., inf.), or *record*/Cro. *snimiti* (v.t., inf.) were provided with the translation which reflected its verbal use in Croatian (all of the listed examples are used in Croatian texts as verbs, never as nouns). There are also instances of entries in the Database for which English loanwords in Croatian are listed as translational equivalents due to their high frequency of use among Croatian speakers. In total, 186 database entries (6.24%) are accompanied by a translational equivalent in Croatian that is an English loanword in origin. If we are to analyze each of the two lists separately, SWEs (150 entries, 8.68% of all SWEs) are more frequently accompanied by English loanwords than MWEs (36 entries, 2.87%) in our Database. English loanwords usually appear

in relation to SWEs denoting a doer of an action (Eng. *babysitter*/Cro. *bejbisiter*, *bejbisiterica*, Eng. *blogger*/Cro. *bloger*, *blogerica*, Eng. *breaker*/Cro. *brejker*, *brejkerica*, Eng. *hater*/Cro. *hejter*, *hejterica*, Eng. *leader*/Cro. *lider*, *liderica*, etc.). As expected, English loanwords also frequently appear among English words from the domains of commerce, economy and business (e.g. Eng. *banner*/Cro. *baner*, Eng. *bestseller*/Cro. *bestseler*, Eng. *budget*/Cro. *budžet*, Eng. *consulting*/ Cro. *konzalting*), popular culture (e.g. Eng. *blockbuster*/Cro. *blokbaster*, Eng. *fake*/Cro. *fejk*, Eng. *fancy*/Cro. *fensi*), sports (e.g. Eng. *bridge*/Cro. *bridž*, Eng. *fitness*/Cro. *fitness*, Eng. *jogging*/Cro. *džoging*) and ICT (e.g. Eng. *cluster*/Cro. *klaster*, Eng. *disc*/Cro. *disk*, Eng. *inch*/Cro. *inč*, Eng. *scart*/ Cro. *skart*). The results in the case of MWEs revealed that out of 36 English expressions only 7 were accompanied by an English loanword as a translational equivalent (e.g. Eng. *spin doctor*/ Cro. *spin doktor*, Eng. *shock horror*/Cro. *šok horor*), whereas in the case of the other 29 MWEs only one phrasal constituent was a loanword from English. Such examples include Eng. *gay friend*/Cro. *gej prijatelj*, Eng. *gala opening*/Cro. *gala otvorenje*, Eng. *ultra clear*/Cro. *ultra čist*, or Eng. *travel blog*/Cro. *putopisni blog*.

Multiple Croatian equivalents were oftentimes available for one and the same meaning (e.g. *bookmark* or *corkage fee*), in which cases all Croatian equivalents were listed along with their respective frequencies. Due to the inflectional nature of the Croatian language, English words referring to people were listed with separate Croatian equivalents where one would denote the male, and the other the female doer (e.g. Eng. *advisor*/ Cro. *savjetnik* (m. nom. sg.), *savjetnica* (f. nom. sg.); Eng. *publisher*/ Cro. *izdavač* (m. nom. sg.), *izdavačica* (f. nom. sg.); Eng. *rookie*/ Cro. *početnik* (n. nom. sg.), *početnica* (f. nom. sg.), *novak* (n. nom. sg.), *novakinja* (f. nom. sg.); etc.). The SWE list includes 65 such entries where both the male and female doer were listed under Croatian equivalents; this figure does not include *rapper*/*rapperica*, *manager*/*managerica*, *teenager*/*teenagerica*, *roller*/*rollerica*, *rocker*/*rockerica*, and *youtuber*/*youtuberica*, which are listed as separate database entries since the expressions denoting female doers in Croatian (*rapperica*, *managerica*, etc.) have been adapted to the Croatian language on the morphological level by the addition of the Croatian inflectional suffix, but have retained English orthographic properties. The MWE list includes 37 such entries where both male and female doers are provided as equivalents (e.g. Eng. *decision maker*/Cro. *donositelj odluka* (m. nom. sg.; f. gen. pl.), *donositeljica odluka* (f. nom. sg.; f. gen. pl.); Eng. *dirty cop*/Cro. *korumpirani policajac* (m. adj. sg.; m. nom. sg.), *korumpirana policajka* (f. adj. sg.; f. nom. sg.); Eng. *gay friend*/Cro. *gej prijatelj* (m. adj. sg.; m. nom. sg.), *gej prijateljica* (f. adj. sg.; f. nom. sg.); Eng. *patient zero*/Cro. *nulti pacijent* (m. adj. sg.; m. nom. sg.), *nulta pacijentica* (f. adj. sg.; f. nom. sg.); etc.). In total, Croatian equivalents for male and female doers are listed separately for 102 entries, which comprises 3.19% of the total number of database entries.

3.5. Semantic categorization

An overview of the database entries from the semantic perspective revealed interesting results in terms of the areas of human activity they originate from, i.e. the specific context in which

they usually appear in the Croatian language. The total count of SWEs and MWEs assigned to each of the 12 categories is listed in Table 4.

Table 4. Representation of the total counts (*n*) and percentages (%) of SWEs and MWEs across the 12 semantic categories

Category	SWEs		MWEs		Total
	n	**%**	***n***	**%**	**%**
PEOPLE	273	15.79	343	27.35	20.66
TECH	351	20.31	131	10.45	16.16
OTHER	291	16.84	141	11.24	14.49
BUSINESS	148	8.56	175	13.96	10.83
HEALTH	165	9.49	158	12.59	10.79
ART	190	10.99	108	8.61	9.99
TOURISM	89	5.15	78	6.22	5.60
FASHION	68	0.04	36	2.87	3.49
FOOD	58	3.36	37	2.95	3.19
SCIENCE	42	2.43	26	2.07	2.28
QUANTITY	41	2.37	16	1.28	1.91
ANT	13	0.75	5	0.39	0.60

Differences between single- and multi-word English expressions have also been observed for the 12 subcategories. The most frequent subcategories on the SWE list were ICT (*n* = 284, 16.44%), SPORT (*n* = 126, 7.29%), and MUSIC (*n* = 82, 4.75%), followed by LANG (*n* = 48, 2.78%), TV (*n* = 41, 2.37%) and COMMERCE (*n* = 39, 2.26%). On the other hand, SPORT (*n* = 104, 8.29%), LANG (*n* = 101, 8.05%), and ICT (*n* = 70, 5.58%) were found to be the most frequent subcategories on the MWE list, followed by COMMERCE (*n* = 38, 3.03%), TV (*n* = 35, 2.79%), and LOC (*n* = 32, 2.55%). In total, the most frequent subcategory in the Database was ICT (*N* = 354, 11.87%), followed by SPORT (*N* = 230, 7.71%) and LANG (*N* = 149, 4.99%).

The highest percentage of database entries was found to belong to the PEOPLE category (20.66%), i.e. they were related to human behaviour and social activity, as well as social phenomena in general. Some of the examples of database entries assigned to this category include: words and expressions related to specific people or groups (e.g. *youtuber* (*RF* = 2.9318) and *youtuberica* (*RF* = 0.4364), *millennials* (*RF* = 0.0544), *hooligan* (*RF* = 0.0970), *homeless people* (*RF* = 0.0052), etc.); words related to human (social) activity (e.g. *crowdfunding* (*RF* = 2.1826),

bullying (*RF* = 1.0180), *mobbing* (*RF* = 2.9945), *dating* (*RF* = 0.4373), etc.), and words related to social phenomena, e.g. activism surrounding human sexuality rights (e.g. *straight* (*RF* = 0.5610), *gay* (*RF* = 24.0839), *queer* (*RF* = 2.8626), *drag queen* (*RF* = 0.2211), etc.). These results could be related to the role of the Internet in today's society, where English is the dominant language. Social networking and the Internet in general have been recognized as activities that facilitate spontaneous vocabulary acquisition (e.g. Godwin-Jones, 2019; Zourou, 2012).

The TECH category was the second most frequently recorded category in the Database (16.16%). The recorded results are not surprising if we consider that the inflow of English words in the last few decades closely follows the growth of the ICT industry, namely the Internet. SWEs like *page* (*RF* = 3.5898; Cro. *stranica*, f. nom. sg.), *memory* (*RF* = 1.9727; Cro. *memorija*, f. nom. sg.), and *domain* (*RF* = 0.3248; Cro. *domena,* f. nom. sg.) are used in Croatian contexts only in reference to ICT, and never to refer to their generic denotations (this is why the Croatian word *memorija* (as in 'computer memory') was used as the translational equivalent for *memory*, and not *sjećanje* ('a memory or the act of remembering', n. nom. sg.), which in Croatian can never be used to refer to the ability of a machine to memorize information, but only to the human capacity to remember). The influence of ICT is also evident in terms of borrowed multi-word units, with English MWEs such as *big data* (*RF* = 0.5395), *cloud computing* (*RF* = 0.4135), and *flat rate* (*RF* = 0.5384) frequently appearing in the Croatian corpora. One possible reason for the frequent use of ICT-related English words could be positive attitudes towards English words, especially in this domain (e.g. Drljača Margić, 2014).

3.6. Per-corpus analysis

A per-corpus analysis of database entries revealed significant variations in frequencies collected for some entries. Table 5 shows the 10 SWE and MWE entries with the highest relative frequencies (*rf*) in each corpus.

Table 5. Database entries with the highest per-corpus frequencies on the SWE and MWE lists

ENGRI					
SWE Entry	**E*af***	**E*rf***	**MWE Entry**	**E*af***	**E*rf***
real	86346	99.3957	**big brother**	9657	11.1165
show	69705	80.2397	**triple(-)double**	2920	3.3613
fan	40273	46.3596	**stand(-)up**	2161	2.4876
summit	28575	32.8936	**fast food**	1755	2.0202
web	27648	31.8265	**fair play**	1606	1.8487
online	25053	28.8393	**plus size**	1531	1.7624
rock	20308	23.3772	**open air**	929	1.0694
jazz	20269	23.3323	**must have**	828	0.9531

ENGRI					
e-mail	19318	22.2376	**rock and roll**	820	0.9439
post	18565	21.3708	**street food**	800	0.9209
hrWac					
SWE Entry	**Haf**	**Hrf**	**MWE Entry**	**Haf**	**Hrf**
web	116672	83.4708	**big brother**	4833	3.4577
blog	112350	80.3787	**red carpet**	3061	2.1899
post	85431	61.1200	**open source**	2949	2.1098
link	79778	57.0757	**fast food**	2532	1.8115
net	57462	41.1101	**made in**	2510	1.7957
e-mail	49698	35.5555	**black carpet**	2466	1.7643
real	46730	33.4321	**off topic**	2129	1.5232
online	43498	31.1198	**stand(-)up**	2055	1.4702
mail	42794	30.6162	**fair play**	2036	1.4566
fan	39346	28.1494	**must have**	1596	1.1418

With further comparison of the two corpora it has been established that some entries appeared in one corpus and never in the other. With regard to SWEs, 28 of them appeared in *hrWaC* and never in *ENGRI*, whereas 14 SWEs were found in *ENGRI* that never appeared in KWIC searches in *hrWaC*. Words like *generally* (H*rf*= 0.47), *chain* (H*rf*= 0.37), and *screencast* (H*rf*= 0.09) were never found in the *ENGRI* corpus, despite the word *generally*, for example, appearing 662 times in *hrWaC*. Similarly, *selfie* (E*rf*= 7.08) and *blockchain* (E*rf*= 1.24) appeared more than 1,000 times in *ENGRI*, but were never found in *hrWac*. Similar results were obtained for MWEs in our Database, with 64 of them never appearing in *ENGRI* (e.g. *mind map* (H*rf*= 0.01), *girls' night out* (H*rf*= 0.01), *critical art* (H*rf*= 0.01), etc.), and 319 MWEs found in *ENGRI*, but never in *hrWaC* (e.g. *ticket point* (E*rf*= 0.25), *ice bucket* (E*rf*= 0.15), etc.). These differences may reflect the differences between the two corpora: while *ENGRI* contains texts collected exclusively from news portals, *hrWaC* also includes texts from blogs, forums, etc. Another possible explanation could be the time period in which the texts were collected. Some words, like *selfie*, could have become more popular after the *hrWaC* corpus had been compiled.

3. Concluding remarks

The focus of research on borrowed words in Croatian has primarily been on loanwords which have undergone adaptation to the recipient language. However, the research outlined in this paper highlights the significance of unadapted English words which can also be found in the Croatian language. Their number and frequency of occurrence in the Croatian corpora suggest they have transcended the boundaries of a simple linguistic phenomenon; a considerable number of English words continue to appear in use despite the fact that acceptable Croatian equivalents

are readily available to language users. This can be taken as evidence corroborating the prestigious status of English among speakers of other languages, as well as proof of its overarching influence on all domains of human activity, especially ICT and popular culture.

The Database, in its current size and scope, presents a valuable addition to language resources in view of open-science policy. Both the Database and the *ENGRI* corpus, created primarily for the purposes of database compilation, are freely available as tools for other researchers whose topics of interest include, but are not limited to, language contact, borrowing process, language prestige, corpus linguistics, or even cognitive processing of foreign words in a recipient language. It serves as a unique tool for the Croatian language, offering a systematic representation of unadapted English words, while also providing insight into the frequency of their use among Croatian language speakers. Furthermore, the model of data representation in the Database provides a foundation for all types of contrastive linguistic research on borrowed lexis, where various factors such as word length, type, or frequency are in focus. Since our data are time-sensitive in nature, our intention is to repeat the compilation process and gather data from texts published after 2020, which would allow us to conduct diachronic studies into the status of English words in Croatian.

ACKNOWLEDGEMENTS

The study outlined in this paper has been supported in part by the Croatian Science Foundation (HRZZ) under project number UIP-2019-04-1576.

References

[Dataset] Bogunović, I., Jelčić Čolakovac, J. & Borucinsky, M. (2022). The database of English words and their Croatian equivalents. figshare. DOI: https://doi.org/10.6084/m9.figshare.20014712.v1

[Dataset] Bogunović, I. & Kučić, M. (2021). *Korpus hrvatskih novinskih portala ENGRI* [Corpus of Croatian news portals ENGRI]. https://urn.nsk.hr/urn:nbn:hr:187:920822.

[Dataset] Bogunović, I., Kučić, M., Ljubešić, N. & Erjavec, T. (2021). Corpus of Croatian news portals ENGRI. Slovenian language resource repository CLARIN.SI. http://hdl.handle.net/11356/1416

[Dataset] Bogunović, I. & Kučić, M. (2022). The database of English words in Croatian.xlsx. figshare. DOI: https://doi.org/10.6084/m9.figshare.20014364.v1

Brdar, I. (2010). *Engleske riječi u jeziku hrvatskih medija* [English words in the language of Croatian media]. *Lahor 10*, 174–189.

Alex, B. (2005). An unsupervised system for identifying English inclusions in German text. In C. Callison-Burch & S. Wan (Eds.), *43. Proceedings of the Annual Meeting of the Association for Computational Linguistics* (pp. 133–138). The University of Michigan. https://dl.acm.org/doi/10.5555/1628960.1628985

Alvarez-Mellado, E. (2020). An annotated corpus of emerging Anglicisms in Spanish newspaper headlines. In *Proceedings of The 4th Workshop on Computational Approaches to Code Switching* (pp. 1–8). European Language Resources Association. https://arxiv.org/abs/2004.02929

Andersen, G. (2012). Semi-automatic approaches to Anglicism detection in Norwegian corpus dana. In C. Furiassi, V. Pulcini & F. R. González (Eds.), *The anglicization of European lexis* (pp. 111–130). John Benjamins. https://doi.org/10.1075/z.174.09

Bogunović, I. & Kučić M. The database of English words in Croatian. Under review.

Bujas, Ž. (2019). *Novi englesko-hrvatski rječnik* [The new English-Croatian dictionary]. Zagreb: Nakladni zavod Globus.

Crystal, D. (2003). *English as a global language* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511486999

Čepon, S. (2017). Anglicizmi v poslovni nomenklaturi turistinih podjetij v Sloveniji. *Revija za ekonomske in poslovne vede* 2, 35–49.

Ćoso, B. & Bogunović, I. (2017). Person perception and language: A case of English words in Croatian. *Language & Communication*, *53*, 25–34. https://doi.org/10.1016/j.langcom.2016.11.001

Drljača, B. (2006). *Anglizmi u ekonomskome nazivlju hrvatskoga jezika i standardnojezična norma* [Anglicisms in the economic terminology of the Croatian language and the standard language norm]. *Fluminensia*, 18(1), 65–85.

Drljača Margić, B. (2014). Contemporary English influence on Croatian: A university students' perspective. In A. Koll-Stobbe & S. Knospe (Eds.), *Language Contact Around the Globe* (Proceedings of the LCTG3 Conference, pp. 73–92). Peter Lang.

Entlová, G. & Mala, E. (2020). The occurrence of anglicisms in the Czech and Slovak lexicons. *Xlinguae, 13*(2), 140–148. https://doi.org/10.18355/XL.2020.13.02.11

Filipović, R. (1990). *Anglicisms in Croatian or Serbian: Origin – development – meaning*. Školska knjiga.

Furiassi, C. & Hofland, K. (2007). The retrieval of false anglicisms in newspaper texts. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years On* (pp. 347–363). Brill/Rodopi. https://doi.org/10.1163/9789401204347_020

Görlach, M. (Ed.). (2002). *An Annotated Bibliography of European Anglicisms*. Oxford University Press. https://doi.org/10.1515/9783484431027.15

Godwin-Jones, R. (2019). Contributing, creating, curating: Digital literacies for language learners, *language learning & technology*, *19*(3), 8–20. https://www.lltjournal.org/item/10125-44427/

Greenall, A. K. (2005). To translate or not to translate: Attitudes to English loanwords in Norwegian. In B. Preisler, A. Fabricius, H. Haberland, S. Kjærbeck & K. Risager (Eds), *The consequences of mobility* (pp. 212–226). Roskilde University.

Halonja, A. & Hudeček, L. (2014). Pokloni mi svoj *selfie* [Give me your *selfie*]. *Hrvatski jezik*, *2*, 26–27.

Hudeček, L. & Mihaljević, M. (2005). *Nacrt za višerazinsku kontrastivnu englesko-hrvatsku analizu* [An outline of a multilevel contrastive Croatian-English analysis]. *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, 31, 107–151. https://hrcak.srce.hr/9381

Jelčić Čolakovac, J. & Borucinsky, M. (2023). In the melting pot of web-crawled texts: The challenges of extracting English words and phrases from Croatian corpora. *International Journal of Applied Linguistics*, *34*(1), 166–182. https://doi.org/10.1111/ijal.12485

Kavgić, A. (2013). Intended communicative effects of using borrowed English vocabulary from the point of view of the addressor: Corpus-based pragmatic analysis of a magazine column. *Jezikoslovlje*, *14*(2–3), 487–499. https://hrcak.srce.hr/112204

Kay, G. (1995). English loanwords in Japanese. *World Englishes*, *14*(1), 67–76. https://doi.org/10.1111/j.1467-971X.1995.tb00340.x

Kilgarriff, A., Rychlý, P., Smrž, P. & Tugwell, D. (2004). Itri-04-08 The Sketch Engine. *Information Technology*, pp. 105–116.

Kučić, M. (2021). Creating a web corpus using GO. In M. Koričić et al. (Eds.), *2021 44th International Convention on Information, Communication and Electronic Technology* (MIPRO) (pp.1676–1678). Croatian Society for Information, Communication and Electronic Technology - MIPRO: Rijeka. DOI: https://doi.org/10.23919/MIPRO52101.2021.9597093

Luján García, C. (2017). Analysis of the presence of Anglicisms in a Spanish internet forum: some terms from the fields of fashion, beauty, and leisure. *Alicante Journal of English Studies*, *30*, 281–305. https://doi.org/10.14198/raei.2017.30.10

Ljubešić, N. & Erjavec, T. (2011). HrWaC and slWac: compiling web corpora for Croatian and Slovene. In I. Habernal & V. Matoušek (Eds.), *Text, speech and dialogue, lecture notes in computer science* (pp. 395–402). Springer.

Ljubešić, N. & Klubička, F. (2016). {bs, hr, sr} wac-web corpora of Bosnian, Croatian and Serbian. In F. Bildhauer & R. Schäfer (Eds.), *Proceedings of the 9th web as corpus workshop (WaC-9)* (pp. 29–35). Association for Computational Linguistics. http://dx.doi.org/10.3115/v1/W14-0405

McKenzie, R. M. (2010). *The social psychology of English as a global language: Attitudes, awareness and identity in the Japanese context*. Springer. https://doi.org/10.1007/978-90-481-8566-5

Međeral, K. (2016). *Jezične bakterije – pomagači ili* štetočine *u jezičnome organizmu?* [Language bacteria – helpers or foes in the language organism?]. *Hrvatski jezik*, 3, 1–10. https://hrcak.srce.hr/171398

Mihaljević Djigunović, J. & Geld, R. (2003). English in Croatia today: Opportunities for incidental vocabulary acquisition. *Studia Romanica et Anglica Zagrabiensia*, *43*, 335–352. https://hrcak.srce.hr/21021

Muhvić-Dimanovski, V. & Skelin Horvat, A. (2006). *O riječima stranoga podrijetla i njihovu nazivlju* [On words of foreign origin and their terminology]. *Filologija*, *44-47*, 203–215. https://hrcak.srce.hr/22242

Muhvić-Dimanovski, V. & Skelin Horvat, A. (2008). Contests and nominations for new words-why are they interesting and what do they show. *Suvremena lingvistika*, *65*(1), 1–26. https://hrcak.srce.hr/25183

Muhvić-Dimanovski, V., Skelin Horvat, A. & Hriberski, D. (2016). *Rječnik neologizama u hrvatskome jeziku* [The dictionary of neologisms in Croatian]. www.rjecnik.neologizam.ffzg.unizg.hr

Nikolić-Hoyt, A. (2005). *Englesko-hrvatski jezično-kulturni dodiri* [English and Croatian in language and cultural contacts]. In D. Stolac, N. Ivanetić & B. Pritchard (Eds.), *Jezik u društvenoj*

interakciji (Zbornik radova sa savjetovanja održanoga 16. i 17. svibnja u Opatiji) (pp. 353–358). Zagreb: Hrvatsko društvo za primijenjenu lingvistiku.

Patekar, J. (2019). *Prihvatljivost prevedenica kao zamjena za anglizme* [The acceptability of loan translations as substitutes for anglicisms]. *Fluminensia*, *31*(2), 143–179. https://doi.org/10.31820/f.31.2.17

Pulcini, V., Furiassi, C. & Gonzales, F. R. (2012). The lexical influence of English on European languages: From words to phraseology. In V. Pulcini, C. Furiassi & F. R. Rodrigues (Eds.), *Anglicization of European lexis* (pp. 1–27). John Benjamins. https://doi.org/10.1075/z.174.03pul

Rüdiger, S. (2018). Mixed feelings: Attitudes towards English loanwords and their use in South Korea. *Open Linguistics*, *4*, 184–198. https://doi.org/10.1515/opli-2018-0010

Serigos, J. R. L. (2017). *Applying corpus and computational methods to loanword research: new approaches to Anglicisms in Spanish*. [Unpublished doctoral thesis]. University of Texas at Austin.

Tadić, M. (2022). *European language equality: Report on the Croatian language*. European Language Equality (ELE): Berlin. https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D1_7__Language_Report_Croatian_.pdf

Tadić, M., D. Brozović-Rončević & Kapetanović, A. (2012). *Hrvatski jezik u digitalnom dobu* [The Croatian language in the digital age]. Springer. https://doi.org/10.1007/978-3-642-30882-6_9

Zourou, K. (2012). On the attractiveness of social media for language learning: a look at the state of the art. *Alsic. Apprentissage Des Langues et Systèmes d'Information et de Communication*, *15*(1). https://doi.org/10.4000/alsic.2436

***

Jasmina Jelčić Čolakovac received her MA degree in English language and History in 2011 at the Faculty of Arts and Sciences in Rijeka. She obtained her PhD degree in Applied Linguistics in 2017 at the University of Ljubljana. Her research interests include English loanwords in Croatian and the processing of metaphoric expressions in bilingual speakers. She has been part of the research team in the newly established Laboratory for Language, Cognition & Neuroscience (LaconLab) since 2020.

Irena Bogunović received her MA degree in English and Croatian languages in 2008 at the Faculty of Arts and Sciences in Rijeka. She obtained her PhD degree in Cognitive Sciences in 2017 at the University of Zagreb. Her research interests include English loanwords in Croatian and their neurocognitive processing by bilingual speakers. She has been acting as the head of the newly established Laboratory for Language, Cognition & Neuroscience (LaconLab) since 2020.

KATARZYNA MROCZYŃSKA[1]
University of Siedlce, Poland
https://orcid.org/0000-0003-0367-1056

DOI: 10.15290/CR.2024.45.2.05

Do *sex* and *gender* go hand in hand? A study of their collocational profiles in EU documents regarding equal treatment of men and women

Abstract. The study of collocations has a long history that goes back to Firth (1957/1968). However, scholarly attention has focused mostly on collocations in general language, with research on this phenomenon within Language for Specialised Purposes (LSP) being a newer and not thoroughly explored line of research[2]. The present article attempts to bridge this gap by looking at the way *sex* and *gender* are employed in the European Union legislation and documents regarding equal treatment of men and women. In particular, the study contrasts and analyses the combinatory potential of *sex* and *gender* as employed in the equal opportunities and non-discrimination regulations and other documents issued by the European Union and its bodies. It also offers a diachronic perspective on how *sex* and *gender* are used in the EU's primary and secondary legislation as well as in guidelines and recommendations. The findings suggest that the two terms in question show completely different collocational profiles and their combinatory potential also varies, with *sex* appearing in a limited number of well-established collocations and *gender* being far more productive and frequent, especially in more recent documents.

Keywords: corpus studies, equal opportunities, legal English, gender, sex, EU regulations

1. Introduction

As this article deals with the corpus of EU legal regulations and official documents regarding equal opportunities, it might be advisable to establish the context and to present the background against which the European non-discrimination law has developed. Prohibition of discrimination on any grounds and equal treatment of men and women are the main principles on which the

1 Address for correspondence: Instytut Językoznawstwa i Literaturoznawstwa, Uniwersytet w Siedlcach, ul. Żytnia 39, 08-110 Siedlce, Poland. E-mail: katarzyna.mroczynska@gmail.com

2 Some insights into legal phraseology, which is the main interest of this study, may be found among others in Góźdź-Roszkowski (2011), Goźdź-Roszkowski and Pontrandolfo (2017), Biel (2012, 2014), Kjaer (1990a, 1990b), Więcławska (2023a, 2023b), and Michta and Mroczyńska (2022).

European Union, originally the European Economic Community, was founded. However, it is worth noting that the body of legal regulations in this area has grown considerably over time and the primary legislation did not always cover the issue in an explicit manner[3].

Protection against discrimination in Europe is provided by both the EU law and the Council of Europe law, with the latter focusing on the European Convention on Human Rights (ECHR). The body of the EU law is largely consistent with the ECHR, the first of the modern human rights treaties that draws from the United Nations Universal Declaration of Human Rights. It sets a legally binding obligation on its members to guarantee a list of human rights to everyone within their jurisdiction, not just citizens. Technically separate and having different origins, structures and objectives, the two systems, i.e. the EU law and ECHR, are to a considerable degree complementary and mutually reinforcing. This is the case despite the fact that the EU itself is not yet a signatory to the ECHR although all 27 member states have ratified the convention. Interestingly, the original treaties of the European Communities did not contain any reference to human rights or their protection. In that time it was believed that the creation of an area of free trade in Europe would not have any impact regarding human rights (Wouters, 2020). Although it turned out quite quickly that the situation was more complex as cases related to alleged breaches of human rights caused by the Community law started to appear in front of the European Court of Justice (ECJ). Consequently, the ECJ developed a set of judge-made laws, the so-called "general principles" of Community Law. Having recognised that its policies could have an impact on human rights, in 2000 the EU and its Member States proclaimed the EU Charter of Fundamental Rights, which included a list of human rights, inspired by the rights contained in the constitutions of the Member States, the ECHR and universal human rights treaties such as the UN Convention on the Rights of the Child. Although in 2000 the Charter was merely a "declaration", it became legally binding in 2009, when the Treaty of Lisbon entered into force. Since then, the EU institutions, like EU Member States, have become legally bound to observe the Charter of Fundamental Rights of the European Union, including its provisions on non-discrimination, but only when they are implementing EU law (Council of Europe: European Court of Human Rights, 2018, pp. 16–22).

All in all, subsequent revisions of the treaties emphasising human dignity, freedom, democracy, equality, the rule of law and respect for human rights led to the Union recognising them as founding values, ones that are not only embedded in the treaties but also mainstreamed into all EU policies and programmes. This shift in perspective is also reflected in the fact that new bodies have been established within the EU such as the European Union Agency for Fundamental Rights (FRA) or the European Institute for Gender Equality (EIGE), their aim being promotion of fundamental rights and equality (Council of Europe: European Court of Human Rights, 2018, pp. 21–23).

3 https://www.europarl.europa.eu/factsheets/en/sheet/59/equality-between-men-and-women

Given the considerable development in the area of equal opportunities, we may expect that the terminology used by the EU legislator to define *sex*/*gender* and equal treatment of women and men in legal regulations has evolved over the years along with the context in which this terminology occurs. We expect these changes to be reflected in the word combinations (collocations[4]) occurring in the corpus of the EU equal opportunities regulations and documents over the years.

At this point, it may also be worth mentioning that the concept of collocation does not only refer to textual statistics, but it reflects a mental representation of the lexicon, as collocations are formed through the cognitive process of priming. As Hoey argues, there are three elementary types of priming: collocation, colligation and semantic preference/association, with the priming of lexical items with collocations in this psychological sense being the foundation of language structure in general (Hoey, 2005, pp. 8–9). In light of these findings, we may assume that knowing how words collocate shows the non-random nature of language (Kilgariff, 2005) and forms an integral part of knowing a language or a genre.

Bearing in mind that collocations reflect a language's conceptual structure, and that a speaker's ability to adhere to collocational conventions demonstrates his/her mastering of the language within a given specific genre, we believe that an analysis of a combinatory potential of words may contribute significantly to the improvement of knowledge of the language and also of the workings of the law as such.

2. Aims and methodology

The purpose of this study was two-fold, namely

1. to analyse the combinatory potential of *sex* and *gender* as employed in the equal opportunities and non-discrimination regulations, as well as in other documents issued by the European Union and its bodies;
2. to analyse how *sex* and *gender* are used diachronically in primary and secondary legislation of the EU as well as in the guidelines and recommendations.

The two aims listed above require the application of a mixed methodology, i.e. corpus linguistics quantitative methods for (1) and mixed quantitative/qualitative methodology of corpus linguistics and discourse studies for (2). We selected a set of legal documents of various genres ranging from the EU primary and secondary legislation (such as the Treaty on European Union,

4 The term *collocation* is credited to Firth (1968), who was the first to spur interest in the habitual company that words keep and draw attention of numerous scholars to this phenomenon. Nowadays, a vast body of literature on this subject is available offering diverse definitions of the term *collocation* as researchers adopt various approaches. An in-depth discussion of research frameworks has already been offered in linguistic literature on numerous occasions and consequently, it is beyond the scope of this study. See among others Sinclair (2004), Kjellmer (1994) or Lehecka (2015) for details of a frequency-based approach, Cowie (1994), Mel'čuk (1998), Hausmann (1997) or Gonzalez-Ray (2002) for a semantic-oriented view, and Siepmann (2005, 2006) for a relatively new, pragmatically-driven approach. An overview of various approaches can be found in Michta and Mroczyńska (2022).

the Treaty on Functioning of EU, the European Convention on Human Rights, the Charter of Fundamental Rights of the European Union, and EU Directives), ancillary documents (e.g. proposals for directives, strategies, recommendations, action plans, a handbook on European non-discrimination law or other guidelines regarding equal opportunities in the EU) to judgements of the Court of Justice referring to equal treatment of men and women. We hope that such an approach will ensure that the language material for the study will be reliable and up-to date[5]. We acknowledge that the corpus compiled in such a manner is relatively small. Yet, the topic it refers to, i.e. equal opportunities, is also a narrow, specialized area. Thus the number of available relevant texts is somewhat limited[6]. After all, any corpus is a kind of compromise between what is planned and desired by the designer and what is possible, for example in terms of available language input or time restrictions (Hunston, 2008, pp. 156–157). Thus, it is worth noting that the corpus we compiled does not lay a claim to being exhaustive and the fact that a collocation does not occur in our corpus does not mean it is definitely invalid in a legal or paralegal text covering equal opportunities. Eventually, our corpus contains 75 documents, 467,472 words, and 594,449 tokens.

The next step was to upload these documents to Sketch Engine, a leading corpus linguistic tool, to allow its investigation. Sketch Engine offers a range of sophisticated functionalities that are useful for retrieving collocations based on selected criteria, including the word sketch, i.e. a condensed description of a word's grammatical and collocational behaviour (Kilgarriff et al., 2014, p. 9). The minimum frequency threshold for retrieving word combinations and potentially identify them as collocations was set at 5 occurrences, meaning that a collocation needs to occur minimum five times to be included in the study. This was done to eliminate potentially invalid word combinations[7]. In the subsequent step, the results produced by Sketch Engine were subject to manual verification. Candidate collocates that upon closer inspection did not act as modifiers were removed from further analysis. The results were then sorted according to a grammatical pattern they appear in. Additionally, sketch difference functionality, which compares the behaviour of two selected words or lemmas, again sorted according to their grammatical patterns, proved extremely helpful in this study. The results obtained with the use of this software functionality were the starting point for the analysis presented in sections 3 and 4 below.

The texts included in the corpus deal with a wide area of equal opportunities as presented in the EU legal and paralegal texts. The frequency list generated for nouns shows that *equality* ranks 6th, *woman* 13th, *man* 37th, whereas *gender* and *sex* were placed in the 31st and 54th position respectively. Our intention was to focus on an analysis of *sex* and *gender* acting as nodes

5 The decision which texts to include in the corpus was based on the summaries of EU legislation in the area of equal opportunities found at `https://eur-lex.europa.eu/EN/legal-content/glossary/equal-opportunities.html`.

6 For a more detailed discussion of building and using small specialised corpora see Koester (2010).

7 That assumption goes in line with Evert (2008, p. 1244), who recommends that a frequency threshold of ≥.5 be applied so as to "weed out potentially spurious collocations".

in our corpus collocations using Sketch Engine. The tool allows for extraction and presentation of search results by different collocational patterns (behaviour) such as (1) premodifier + noun, (2) noun + noun, (3) verb + noun, (4) noun + verb. (5) preposition + noun, (6) noun + preposition (cf. Hausmann, 1989).

The patterns above may be analysed in pairs due to their structural similarity. For example, in pattern (1) above a modifier may be an adjective, a noun or a participle whereas pattern (2) allows for modifications using a noun only (Michta & Mroczyńska, 2022, p. 40). The results obtained in the Sketch Engine search showed that the largest number of collocates may be found with *sex* and *gender* acting as modifiers for nouns, i.e. as in (1) and (2) above. The remaining patterns were only scarcely identified, with software often yielding just a couple of possible word combinations. That is why pattern (2) was the first candidate for more in-depth research.

For the purpose of this comparative analysis, we used Sketch Engine word sketch difference function, which makes it possible to juxtapose collocations of two selected lemmas/words. The list of possible collocates returned by the software shows that the collocability of both words does not overlap, i.e. *sex* will modify a different set of nouns than *gender* will.

3. Combinatory potential of *sex* and *gender* in the EU texts on equal opportunities

The online *Merriam-Webster Dictionary* states that both *sex* and *gender* are well-established words in the English language and their history dates back to the 14th century. *Gender*, deriving from the Latin word *genus* and the Old French *gendre* (Corbett, 1991, p. 1), was used in English to refer primarily to grammatical gender (Hockett, 1958, p. 231). In the 15th century, the meaning of *gender* expanded to include what *sex* had referred to since 14th century, i.e. either of the two primary biological forms of species. Though the online *Merriam-Webster Dictionary* claims that initially *sex* and *gender* were used interchangeably to refer to one of the two primary biological forms of species – male or female, Gries, Slocum and Solan (Brief for *Amici Curiae*, 2019, p. 23) found that in American English *gender* was almost exclusively used to refer to a grammatical category and it was extremely uncommon to use it outside this meaning until the 1960s. Though intertwined, the usage of *sex* and *gender* has evolved and the words have gained new meanings. In the 20th century, *sex* acquired the 'sexual intercourse' meaning, which become the most frequently used, whereas *gender* "gained a meaning referring to the behavioural, cultural, or psychological traits typically associated with one sex, as in *gender roles*".

In this study, we will analyse the combinatory potential of *sex* and *gender*, focusing only on their meanings referring to being male, female or neutral. As we can infer from the definitions provided above, the terms deal with the issue from different angles – the biological or psychological and socio-cultural ones respectively – and consequently they refer to different concepts. The analysis of other meanings of *sex* and *gender* is beyond the scope of our study.

Let us start with a brief overview of definitions of the two terms that may be found in reference books such as dictionaries and glossaries. We consulted selected general and legal dictionaries

as well as glossaries, both British and American ones. The next step was to analyse definitions of the two words in question provided in the reference sources. It appears that they tend to present *sex* as being a biological feature, whereas *gender* rather as a socio-cultural concept and/or a collection of psychological traits. We may notice that despite being different the terms are connected. To shed some light on this quite complex issue, below we offer a compilation of definitions culled from selected sources. We start by providing definitions that may be found in English dictionaries and next move on to those included in glossaries and articles devised by international institutions such as the EU Council, WHO or the UN agendas

Table 1. Selected definitions of the terms *sex* and *gender*

Source	*Sex*	*Gender*
Cambridge Dictionary (online)	1a the physical state of being either male, female, or intersex 1b all males considered as a group, or all females considered as a group	1a a group of people in society who share particular qualities or ways of behaving which that society associates with being male, female, or another identity 1b the condition of being a member of a group of people in a society who share particular qualities or ways of behaving which that society associates with being male, female, or another identity 1c used to refer to the condition of being physically male, female, or intersex (= having a body that has both male and female characteristics)
Collins English Dictionary (online)	1 The two **sexes** are the two groups, male and female, into which people and animals are divided according to the function they have in producing young 2 The **sex** of a person or animal is their characteristic of being either male or female.	1 **Gender** is the state of being male or female in relation to the social and cultural roles that are considered appropriate for men and women 2 You can use **gender** to refer to one of a range of identities that includes female, male, a combination of both, and neither 3 Some people refer to the fact that a person is male or female as his or her **gender**
Merriam Webster	1a either of the two major forms of individuals that occur in many species and that are distinguished respectively as female or male especially on the basis of their reproductive organs and structures 1b the sum of the structural, functional, and sometimes behavioral characteristics of organisms that distinguish males and females 1c the state of being male or female 1d males or females considered as a group	2a sex as 1a 2b the behavioral, cultural, or psychological traits typically associated with one sex
The Law Dictionary	The distinction between male and female; or the property or character by which an animal is male or female.	Defined difference between men and women based on culturally and socially constructed mores, politics, and affairs. Time and location give rise to a variety of local definitions. Contrasts to what is defined as the biological sex of a living creature.

Source	*Sex*	*Gender*
Gender Equality Glossary	Sex refers to the biological characteristics that define humans as female or male. While these sets of biological characteristics are not mutually exclusive, as there are individuals who possess both, they tend to differentiate humans as males and females.	Article 3C of the Istanbul Convention: "Gender shall mean the socially constructed roles, behaviours, activities and attributes that a given society considers appropriate for women and men" Other definitions of "gender": – Gender refers to the social attributes and opportunities associated with being male and female and the relationships between women and men and girls and boys, as well as the relations between women and those between men. These attributes, opportunities and relationships are socially constructed and are learned through socialization processes. They are context/ time-specific and changeable. Gender determines what is expected, allowed and valued in a women or a man in a given context. In most societies there are differences and inequalities between women and men in responsibilities assigned, activities undertaken, access to and control over resources, as well as decision-making opportunities. Gender is part of the broader socio-cultural context. Other important criteria for socio-cultural analysis include class, race, poverty level, ethnic group and age (UN Women). – Gender is a concept that refers to the social differences between women and men that have been learned are changeable over time and have wide variations both within and between cultures (European Commission).
Gender Equality Glossary: terms and concepts	Refers to the biological and physiological reality of being males or females.	A social and cultural construct, which distinguishes differences in the attributes of men and women, girls and boys, and accordingly refers to the roles and responsibilities of men and women. Gender-based roles and other attributes, therefore, change over time and vary with different cultural contexts. The concept of gender includes the expectations held about the characteristics, aptitudes and likely behaviours of both women and men (femininity and masculinity). This concept is useful in analyzing how commonly shared practices legitimize discrepancies between sexes.
WHO (2024)	refers to the different biological and physiological characteristics of females, males and intersex persons, such as chromosomes, hormones and reproductive organs	refers to the characteristics of women, men, girls and boys that are socially constructed

As we can see in Table 1, most references offer a more concise definition of *sex*, whereas the concept of *gender* frequently requires a more extensive and elaborate explanation. The point may be that the meaning of the former seems to be well-established while the latter is a relatively new concept in the public discourse, thus requiring more in-depth explanation[8].

Interestingly, specialised dictionaries of law, at least those we consulted, i.e. British dictionaries including Jowitt's *Dictionary of English Law* (2010), Osborn's *Concise Law Dictionary* (Woodley, 2013) or *Oxford Dictionary of Law* (Law, 2015), as well as American ones such as *Black's Dictionary of Law* (Garner, 2019), *Wex*, *The People's Law Dictionary* (Hill & Thompson Hill, 2002), do not provide definitions of *sex* and *gender.* However, they do include terms containing *sex* or *gender* as modifiers, e.g. *sex discrimination*, *sex change*, *gender reassignment*, *gender pay gap* or *gender bias*[9]. That would imply that *sex* and *gender* are not treated as terminological units in legal English as they are not presented in separate entries.

Having reviewed what reference books offer, we may move on to the comparative analysis of the corpus of texts regarding equal opportunities in the EU. As mentioned in the Aims and methodology section, the preliminary findings generated with the Word Sketch functionality revealed that the combinatory potential of the two words in question concentrated in the noun + noun or premodifier + noun category. In the other categories, the identified word combinations were either infrequent or invalid at times, e.g. *race* was listed as a collocate of *sex* but after a closer investigation it appeared that the only word combination it featured was lists such as […] *sex*, *race*, *colour*, *language, religion* […]. Therefore, the decision was made to focus on collocations where *sex* and *gender* appear as modifiers of other nouns. To facilitate the study, we used the Sketch Engine Word Sketch Difference function, which makes it possible to juxtapose collocations with two selected lemmas/words, in this case *sex* and *gender* respectively. The list of possible collocates that the software returned shows that the collocability of both words does not overlap, i.e. *sex* will modify a different set of nouns than *gender* will. What is more, the results prove that *gender* has a much greater combinatory potential appearing in a wide range of collocations whereas *sex* appears only in two collocations, namely *sex characteristics* and *sex discrimination*, the latter actually being a well-established term featuring in most dictionaries[10].

When it comes to collocates of *gender*, the most frequently appearing one was *equality* (*gender equality* with a frequency of 210), followed by *gap* (52), *identity* (45) and *balance* (43) each one of them occurring in the corpus not nearly as frequently as *gender equality*. The list generated by

8 Developing from and alongside the Women's Studies and feminist movements of 1960s and 1970s, Gender Studies gained popularity in Western universities in 1990s (see among others Wiegman, 2002; Halberstam, 2014).

9 An interesting comparative study in the use of selected collocations in general and specialised legal corpora may be found in Michta (2022).

10 That is in line with what some researchers point out, namely the fact that modifier + noun combinations may cover not only collocations but also terms. See among others Bergenholtz and Tarp (1994), Michta et al. (2009), L'Homme and Azoulay (2020). Distinguishing between collocations and terms may constitute an interesting line of research though it is not the main focus of this study.

Sketch Engine functionality also includes *gender stereotype* (24), *gender inequality* (21), *gender mainstreaming* (11), *gender perspective* (11), *gender bias* (7), *gender role* (7), *gender expression* (6), *gender dimension* (6), and *gender impact* (6). There are also two collocations referring to legal and medical procedures, namely *gender reassignment* (25) and *gender reassignment surgery* (16). The word *gender* also appears as a modifier in the titles of documents as in *Gender Directive* (with *gender directive* appearing as an alternative spelling variant – total frequency of 45), or *Gender Goods and Services Directive* (21), and *gender strategy* (with *Gender Strategy* appearing as an alternative spelling variant with a total frequency of 19).

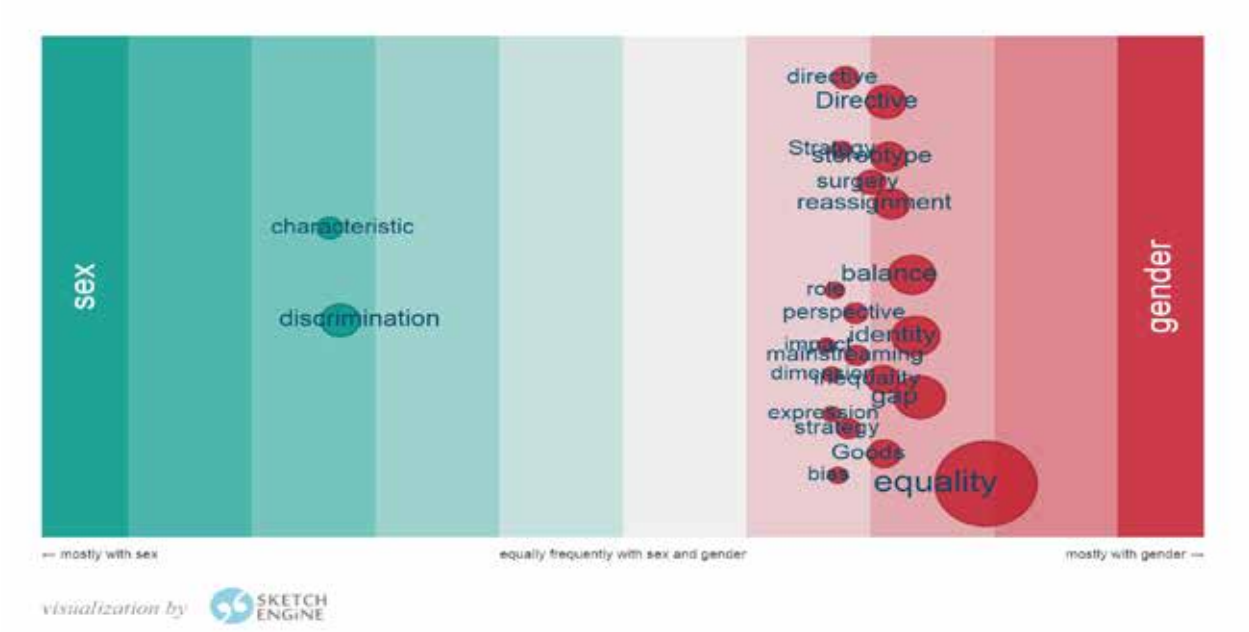

Figure 1. Sketch Engine visualisation of word combinations with *sex* and *gender* acting as modifiers

As we can see from the above analysis, the retrieved collocations vary greatly in their frequency, with *gender equality* being by far the most common word combination. Interestingly, *gender equality* appears in the corpus ten-times more frequently than its antonym, i.e. *gender inequality*, which shows the frequency of only 21. This may be connected with the EU institutional approach, reflected in the language used in the documents, of enforcing positive behaviour rather than punishing or stigmatising the wrong one.

At this point it may be worth mentioning that the analysis of collocations is a useful tool allowing discrimination between near-synonyms, with collocations being a mark of meaning difference (Sinclair, 1991, p. 170), and the findings of this study seem to prove it. The fact that the *sex* and *gender* appear as modifiers in completely different word combinations, as we can see in Figure 1 above, may imply that the words should not be seen as synonyms, at least in

the area of specialist legal language[11]. Generally, in the legal language and legal texts authors should follow the "principles of semantic accuracy and language consistency", which are key to avoiding ambiguity and misunderstandings (Jopek-Bosiacka, 2011, p. 16). Having said that, it may be worth noting that using synonymy in legal contexts is rather unwelcome though not absent from legal texts (cf. Góźdź-Roszkowski, 2013; Matulewska, 2016; Rzepkowska, 2023).

Approaches to recognising and classifying synonyms across linguistic literature are numerous, though the researchers' views may vary they do share one feature, namely interchangeability/substitution which seems to be one of the persistent criteria in identifying potential synonyms (Crystal, 2003, p. 450; Lewandowska-Tomaszczyk, 1990)[12]. The analysis of our corpus reveals that the two words in question are not interchangeable. That is why it may be argued that *sex* and *gender* are not cognitive synonyms, though may constitute plesionyms. Cruse (1986, p. 285) introduces plesionyms as a separate category, different from cognitive synonyms, and describes them as words that cannot mutually entail; that is to say, there seems to be some overlap in their meaning but they cannot be used interchangeably. That is exactly the case of *sex* and *gender* as analysed in the EU legal context regarding equal treatment of men and women.

4. A diachronic terminological shift in the EU legislation and proposals

The documents included in the corpus were published over the period of about 70 years. The earliest is primary legislation which dates back to 1950s (e.g. the European Convention on Human Rights came into force on 3rd September 1953, the Treaty on the EU of 2009 is based on the Treaty of Rome establishing the EEC in 1957, the Treaty on the Functioning of the EU became effective in 2016, and the Charter of Fundamental Rights of the EU was proclaimed in 2000 and given legal effect in 2009). The secondary legislation, i.e. directives, cover the time span from 1979 to 2022, whereas the date of publication of documents covering recommendations and guidelines in the area of equal opportunities ranges from 2014 to 2023.

While conducting this analysis, we noticed that the extracted collocations tend to fall into one or more of the three categories, i.e. those appearing in primary legislation, in secondary legislation, or in proposals and recommendations. This breakdown might indicate that the wording the EU bodies apply to refer to equal opportunities evolves and that in turn may be a reflection of the legislator's or the societies' changing needs and/or perspective in this area of regulation[13].

---

11 Studies discussing synonymy in general language include Lyons (1981), and Cruse (1986, 2000); the legal context is presented in Matilla (2006), Góźdź-Roszkowski (2013), Matulewska (2016), Klabal (2019), and Rzepkowska (2023).

12 See Landau (2001, p. 137) for the treatment of synonymy in dictionaries.

13 An overview of the use of words *sex* and *gender* in the EU primary and secondary legislation from a legal system perspective may be found in Siekera (2022).

Working with the corpus data, we notice a shift in the wording the EU uses when dealing with the issue, i.e. from a non-discrimination approach (which may be found mostly in older, primary legislation) to promoting equal opportunities (in more recent, secondary legislation and recommendations). This change may be due to the fact that either basic non-discrimination issues have been regulated sufficiently and the EU bodies may move on to devising more refined regulations regarding promotion of equal opportunities and not just combating inequalities. The shift may also reflect the ambition of the anti-discrimination laws which is not just to change behaviour but to change cognitions about and emotions towards stereotyped groups (cf. Allport, 1979). Legal regulations can accomplish their goals directly, through fear of sanctions or desire for rewards. But they can also do so indirectly, by changing attitudes about the regulated behaviours. To this end, the law may implicitly or explicitly incorporate findings from psychological science which help understand how individuals think, feel, and make decisions (Nadler & Mueller, 2017, pp. 124–125; see also Bilz & Nadler, 2014). That leads us to the role that the language used in regulations may play in this process.

The language we use may have an effect on the way we perceive the world around us (cf. Wharf, 1956; Zlatev & Blomberg, 2015). Therefore, the collocation analysis carried out based on the authentic corpus material may offer an insight into how words and phrases are used and shed some light on associations that particular words or phrases may carry for language users (Taylor, 2021, p. 572). Baker (2006, p. 13) points out at an interesting application of a collocation analysis in discourse studies, i.e. "collocates may be helpful at revealing how meaning is acquired through repeated uses of language, as certain concepts become inextricably linked over time". What is more, as Bogetić argues this kind of analysis may also be employed as a means to understand ideology as lexical co-occurrence of words helps uncover "a complex web of identities discourses and social representations in communities" (Bogetić, 2013, p. 334).

The kind of analysis we are going to present in this section combines the two areas of language research, i.e. corpus linguistics (with its quantitative approach) and critical discourse analysis (relying on qualitative methods), crossing the traditional clear-cut dividing line between research methodologies.

When it comes to the first category of documents, i.e. EU primary legislation, the terms *sex* and *equality between men and women* are used. We did not find any collocations having *sex* or *gender* as a premodifier, i.e. in the collocational pattern that was the focus of this study. The Treaty on European Union as well as the Treaty on the Functioning of the European Union point out that the activities of the EU shall aim at eliminating inequalities, combating discrimination and promoting equality between men and women. The non-discrimination on the grounds of sex is also included in the Charter of Fundamental Rights, which also advocates applying measures providing for a specific advantage of the *under-represented sex*. The legislator uses the term *sex* and collocates of this word when referring to the area of equal opportunities. The term *gender* does not appear in these documents. This may be due to the fact that primary legislation was drafted at a time when the notion of *gender* as opposed to *sex* was non-existent

in legal regulations[14]. However, the analysis of Directives, i.e. the secondary legislation, shows a much more varied collocational landscape.

The seven directives which create the legal framework for the implementation of the EU funding principle of non-discrimination show an evolution in the terms they employ when dealing with equal treatment of men and women. We will analyse selected documents in chronological order. First, we put under scrutiny the earliest directive regulating this area, i.e. the Council Directive 79/7/EEC on the progressive implementation of the principle of equal treatment of men and women in matters of social security. It defines equal treatment as non-discrimination, direct or indirect, based on *sex*. In Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, the term *sex* does not appear at all whereas *gender* is used once in Article 17 (2) which states that the principle of *gender mainstreaming* should be applied when preparing an assessment of the impact of the measures taken on women and men. In EU documents, gender mainstreaming is understood as means to an end of realising gender equality, involving "the integration of a gender perspective into the preparation, design, implementation, monitoring and evaluation of policies, regulatory measures and spending programmes, with a view to promoting equality between women and men, and combating discrimination that involves" (European Institute for Gender Equality). Council Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation goes along similar lines as the above-mentioned Directive 79/7/EEC referring to discrimination on the grounds of *sex* when addressing the issue of fixing for occupational social security schemes of ages for admission or entitlement to retirement or invalidity benefits. The very Directive also mentions the principle of *gender mainstreaming*. Another document from the legal framework of equal opportunities for men and women worth mentioning here is Gender Goods and Services Directive (2004/113/EC) which uses the concept of *gender equality* when discussing equal treatment of men and women in their access to and supply of goods and services. The document also contains references to the strategy on *gender equality* and recommends promotion of *gender equality*. Rather predictably, the term *sex* is used when addressing the issue of *discrimination based on sex* or combating *sex discrimination*. It also makes a reference to *members of one sex* or *a person of one sex or the other*, which implies a binary notion of sex as used in this document. Interestingly, both *sex* and *gender* may be found in Gender Equality Directive 2006/54/EC. Again, the term *sex* appears in the contexts involving both direct and indirect discrimination, whereas when it comes to the term *gender* it appears when referring to the concept of *gender mainstreaming*, discrimination arising from *gender reassignment*, *gender-based wage differentials*, *gender segregation on the labour market* or the

14 The origin of the concept of *gender* and the term *gender identity* goes back to American psychiatric research of the 1950s and 1960s, John William Money (1955), a sexologist, and Robert Stoller (1964, 1968), a psychiatrist, respectively. For details of the history of defining *sex* and *gender*, the reader is referred to Schiappa (2022, p. 15-32).

EU body dealing with equal opportunities, namely "the future European Institute for Gender Equality". Finally, Directive 2010/41/EU on the application of the principle of equal treatment between men and women engaged in an activity in a self-employment capacity also uses *sex* and *gender* in quite a similar way as Gender Equality Directive, i.e. *sex* for discrimination-related provisions and *gender* to refer to *gender equality* and *gender mainstreaming*. There is one more directive addressing equal opportunities issues worth including in this overview, namely Directive 2022/2381/EC on improving the gender balance among directors of listed companies and related measures. Although the regulation covers a relatively narrow scope, that is gender balance on boards of directors of listed companies, it deals with the topic in a very meticulous manner. Apart from employing well-established collocations such as *sex discrimination* and *discrimination based on sex*, the document refers to *a person of the other sex* and on numerous occasions makes provisions regarding *the under-represented sex*. When it comes to *gender*, apart from a frequently occurring collocation *gender equality*, the corpus analysis yields such phrases as *to foster gender-balanced decision making*, *to close the gender (pay) gap* and to *achieve a gender-balanced representation* among top management positions. The fact that a directive covering this issue was adopted quite recently may suggest that sex discrimination is still an issue – even at top managerial levels in large organisations – and that the glass ceiling still exists.

Table 2. Collocations with *sex* and *gender* as a premodifier retrieved from the EU directives

Premodifier	Collocation	Frequency
sex	sex discrimination	6
gender	gender balance	27
	gender equality	24
	gender gap	8
	gender mainstreaming	2
	gender reassignment	1
	gender segregation	1

All in all, in the analysis of secondary legislation a pattern emerges where the use of *sex* is limited to the context of discrimination and the fact of being male or female, whereas *gender* is used in the context of promoting equal opportunities for men and women (*gender equality*, *gender mainstreaming*) or pointing out existing inequalities (*gender-based wage differentials*, *gender segregation on the labour market*). We can also see that the number of collocations with *sex* and *gender* is higher than in the first category of documents analysed. This may be explained by the fact that the progressive EU policies in this area have been gaining their momentum, which is also reflected in the expansion of the lexicon.

Last but not least, the third category of texts, which includes documents outlining the directions or making recommendations for future actions such as strategies, strategic framework, Commission recommendations, proposals for directives when addressing equal opportunities issues, make reference to *gender* rather than *sex.* These documents often offer detailed descriptions of more recently established contexts and concepts in the area of non-discrimination, which have not been reflected in the legislation yet, or the recommendations may be paving a way to some new approaches.

Rather unsurprisingly, and in line with our previous findings, the majority of collocations culled from this group of documents are based on *gender* and not *sex*. Thus, we retrieve word combinations such as *gender equality*, *gender identity* (as in *(non-) discrimination on the ground of gender identity*), *gender gap also pay gap*, *gender employment gap* or *gender care gap* (in phrases such as *to close/tackle a gender pay gap)*, *gender stereotype* (often in verb phrases such as *to challenge/ combat/ debunk/ address or recognise gender stereotypes*), *gender balance* (e.g. *to improve/achieve/ensure gender balance*), *gender inequality* (e.g. *to eliminate/identify/reduce gender inequality*), *gender reassignment or gender reassignment surgery* (e.g. *to undergo a gender reassignment surgery, discrimination based on/arising from gender reassignment*), *gender perspective* (e.g. *to adopt/mainstream/integrate/ the gender perspective*, *from a gender perspective*), *gender strategy* and *gender equality strategy*, *gender mainstreaming* (e.g. *to strengthen/enhance/ improve gender mainstreaming, the principle of gender mainstreaming*), *gender bias* (e.g. *to address/reveal/challenge a gender bias*), *gender role* (e.g. *to strengthen/reinforce (traditional) gender roles, a traditional distribution of gender roles*), *gender expression* (e.g. *discrimination based on gender expression*), *gender dimension* (e.g. *to have/address/integrate a gender dimension*), and *gender impact* (e.g. *to consider/focus on/ look at (the) gender impact of sth*)[15].

15 The collocations are listed by their frequency in the corpus, from the most to the least frequent ones.

Table 3. Collocations with *sex* and *gender* as a premodifier retrieved from the EU guidelines and recommendations

Premodifier	Collocation	Frequency
sex	sex discrimination	21
	sex characteristics	13
gender	gender equality	185
	gender identity	45
	gender gap	44
	gender stereotype	24
	gender inequality	20
	gender reassignment surgery	14
	gender balance	12
	gender reassignment	12
	gender perspective	11
	gender strategy	11
	gender mainstreaming	9
	gender bias	7
	gender role	7
	gender expression	6
	gender dimension	6
	gender impact	6

As already mentioned, in this group of documents, and in the whole corpus, collocations with *gender* acting as a modifier outweigh those modified by *sex*. Still, *sex characteristics* appears frequently alongside *sexual orientation*, *gender identity*, *gender expression* as potential grounds of discrimination. The fact that recommendations and strategies introduce a number of new word combinations (concepts) absent from primary and secondary legislation may reflect the EU strive to ensure equal treatment for all its citizens in various aspects of their lives and extend non-discrimination protection, which in turn is linked to the changing needs and lifestyles.

The type and frequency of word combinations retrieved from the EU primary and secondary legislation as well from guidelines and recommendations analysed against their timeline may suggest that the language used in the documents has evolved. It seems that the focus has shifted from biological *sex* as potential grounds for discrimination to *gender* as a wider category embracing a number of social and cultural issues the EU member states and the EU bodies may

need to address in order to ensure equal treatment of men and women. The EU seems to be a place where women enjoy a relatively strong position compared to many other parts of the world, where their rights are often challenged. However, even within the EU there have been some setbacks and persistent difficulties. Therefore, progress is still required in the political, economic and social fields to achieve true gender equality in all of these areas (Buzmaniuk, 2023, p. 1).

5. Findings and conclusions

The analysis reported in the previous two sections explored the frequency and combinatory potential of words *sex* and *gender* as used in the EU documents dealing with equal treatment of men and women. The main findings can be summarised as follows:

1. the fact that collocates that feature *sex* and *gender* do not overlap, not only in the modifier +term category, but in all the other categories retrieved by Sketch Engine, suggests that the words are not synonyms but occupy a different place in the semantic space, and consequently their collocational profiles vary. Clearly not being synonyms, they may be considered plesionyms though as there is some overlap in their meaning but they cannot be used interchangeably. Collocations a term may enter are a mark of the meaning difference, and the findings of this study seem to prove it. This observation may have implications for discriminating between near-synonyms and for teaching legal English genre (see Yevchuk (2021) for her study of Estonian learners).
2. our computational analysis of the corpus shows that the frequency of the noun *gender* surpasses that of *sex*, the former occurring 806 times and the latter 500 times, with *gender* constituting 61.7% of analysed corpus occurrences of the two terms and *sex* accounting for 38.3%. Thus, we can see that the frequency of *gender* in the analysed documents is 61.2% higher than that of *sex*. This difference is also reflected in their respective collocates. However, we can see that the relationship is not linear. The number of word combinations with *gender* as a modifier shows much greater prevalence since *sex* is retrieved only in two collocations, namely *sex discrimination* and *sex characteristic* while *gender* features in dozens of word combinations as listed in sections 3 and 4 of this paper.
3. the frequency of the two words has evolved over time, and so has their combinatory potential. Whereas the combinatory potential of the word *sex* remained rather stable, the word *gender* has entered the lexicon with great force and the number of new phrases that it features seems to be constantly growing with new collocations appearing in the corpus. The arrival of relatively new phrases may be observed in particular in such documents as strategies, recommendations or directive proposals. That may be due to regulatory needs and changes in societies in the member states, e.g. changing values/beliefs and attitudes regarding equal treatment of men and women. The more recent legislation has introduced new concepts that refer rather to social psychological sphere rather than purely biological one. That may explain why *gender* and its word combinations have proliferated and why

they clearly outnumber combinations featuring *sex* as a modifier. The change in regulatory approach may be reflected in the language used in the area of equal treatment since legislation is a tool used to modify some behaviours.

All in all, we are aware of the fact that the study was limited to specialised legal English as used in the EU texts in the area of equal opportunities for men and women and focused solely on the modifier + noun type of collocations. The collocations retrieved from a legal corpus and a general one may differ. Thus, this issue may necessitate conducting further research and the findings provided may also have implications for the development of collocation-centred language teaching resources. The context, a general or specialised one, may call specifically for language material compiled from a general or specialised corpora, respectively (cf. Michta, 2022, p. 90).

Still, a tentative observation can be made that the collocations occurring in the small specialized corpus we analysed reflect the changing values and beliefs in societies which are reflected in the nature of the language used. It would be in line with the view that the legal language and consequently permitted word combinations are inextricably intertwined with a particular legal system (Kjær, 2007, p. 508), in this case with the European Union legal regulation system.

References

Allport, G. W. (1979). *The nature of prejudice: 25th anniversary edition*. Basic Books.

Baker, P. (2006). *Using corpora in discourse analysis*. Continuum.

Bergenholtz, H. & Tarp, S. (1994). Mehrworttermini und Kollokationen in Fachwörterbüchern. In B. Schaeder & H. Bergenholtz (Eds.), *Fachlexikographie: Fachwissen und seine Repräsentation in Wörterbüchern* (pp. 385–419). Gunter Narr Verlag.

Biel, Ł. (2012). Areas of similarity and difference in legal phraseology: collocations of key terms in UK and Polish company law. In A. Pamies, J. M. Pazos Bretaña & L. Luque Nadal (Eds.), *Phraseology and discourse: Cross linguistic and corpus-based approaches* (pp. 225–233). Schneider Verlag.

Biel, Ł. (2014). Phraseology in legal translation: a Corpus-based analysis of textual mapping in EU law. In L. Cheng, K. Kui Sin & A. Wagner (Eds.), *Ashgate handbook of legal translation* (pp. 178-192). Ashgate Publishing.

Bilz, K. & Nadler, J. (2014). Law, moral attitudes and behavioural change. In E. Zamir & D. Teichman (Eds.), *Oxford handbook of behavioural economics and the law* (pp. 241–267). Oxford University Press.

Bogetić, K. (2013). Normal straight gays: Lexical collocations and ideologies of masculinity in personal ads of Serbian gay teenagers. *Gender & Language, 7*(3), 333–367.

Brief for *Amici Curiae* Corpus-Linguistics Scholars Professors Brian Slocum, Stefan Th. Gries, and Lawrence Solan supporting employees, Bostock v. Clayton County, Altitude Express v. Zarda & Moore, and Harris Funeral Homes v. EEOC. (2019).

Buzmaniuk, S. (2023). Gender equality in Europe: a still imperfect model in the world. *European Issues, 659*, 1–7. Foundation Robert Schuman Policy Paper.

Corbett, G. (1991). *Gender*. Cambridge University Press.

Cowie, A. (1994). Phraseology. In R.E. Asher & J.M.Y. Simpson (Eds.), *The encyclopedia of language and linguistics* (pp. 3168–3169). Pergamon Press.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.

Cruse, D. A. (2000). *Meaning in language: An introduction to semantics and pragmatics*. Oxford University Press.

Crystal, D. (2003). *A dictionary of linguistics and phonetics* (5th edition). Blackwell Publishing.

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook* (pp. 1212–1248). De Gruyter.

Firth, J. R. (1957/1968). A synopsis of linguistic theory, 1930–1955. In F.R. Palmer (1968) (Ed.), *Selected Papers of J.R. Firth 1952–1959* (pp. 168–205). Indiana University Press.

Garner, B. A. (2019). Editor in chief. *Black's dictionary of law* (11th edition). Thomson West.

Gonzalez-Ray, I. (2002). *La phraséologie du français*. Presses Universitaires du Mirail.

Goźdź-Roszkowski, S. (2011). *Patterns of linguistic variation in American legal English: A corpus-based study*. Peter Lang.

Goźdź-Roszkowski, S. (2013). Exploring near-synonymous terms in legal language. A corpus-based, phraseological perspective. *Linguistica Antverpiensia, New Series – Themes in Translation Studies, 12*, 94–109.

Goźdź-Roszkowski, S., & Pontrandolfo, G. (Eds.). (2017). *Phraseology in legal and institutional settings: a corpus-based interdisciplinary perspective*. Routledge.

Halberstam, J. (2014). Gender. In B. Burgett & G. Hendler (Eds.), *Keywords for American Cultural Studies* (2nd edition, pp. 116–18). NYU Press. http://www.jstor.org/stable/j.ctt1287j69.33.

Hausmann, F. J. (1989). Le dictionnaire de collocations. In F. J. Hausmann, O. Reichmann, H. E. Wiegand & L. Zgusta (Eds.), *Wörterbücher: ein internationales Handbuch zur Lexicographie* (pp. 1010–1019). De Gruyter.

Hausmann, F. J. (1997). Tout est idiomatique dans les langues. In M. Martins-Baltar (Ed.), *La locution entre langue et usages* (pp. 227–290). ENS Éditions.

Hill, G. N. & Thompson Hill, K. (2002). *The People's Law Dictionary. Taking the Mystery Out of Legal Language*. New York: MJF Books. Also available online at https://dictionary.law.com/default.aspx?review=true

Hockett, Ch. F. (1958). *A course in modern linguistics*. Macmillan.

Hoey, M. (2005). *Lexical priming: A new theory of words and language*. Routledge.

Hunston, S. (2008). Collocation strategies and design decisions. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook* (pp. 154–168). De Gruyter.

Jopek-Bosiacka, A. (2011). Defining law terms: A cross cultural perspective. *Research in Language. Special Issue on Legal Terminology: Approaches and Applications*, *9*.1, 9–29.

Jowitt's dictionary of English law (3rd edition). 2010. Sweet and Maxwell.

Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*, *1/2,* 263–276.

Kilgarriff, A., Baisa V., Bušta J., Jakubíček M. & Kovář V., Michelfeit J., Rychlý P. & Suchomel V. (2014). The Sketch Engine: ten years on. *Lexicography, 1*, 7–36.

Kjaer, A. L. (1990a). Context-conditioned word combinations in legal language. *Terminology Science & Research* 1/1–2, 21–32.

Kjaer, A. L. (1990b). Phraseology research – State-of-the-art. *Terminology Science & Research, 1*/1–2, 3–20.

Kjær, A. L. (2007). Phrasemes in legal texts. In H. Burger, D. Dobrovol'skij, P. Kühn & N. R. Noerrick (Eds.), *Phraseologie/Phraseology: Ein internationales Handbuch zeitgenössischer Forschung / An international handbook of contemporary research* (pp. 506–516). de Gruyter.

Kjellmer, G. (1994). *A dictionary of English collocations*. Clarendon Press.

Klabal, O. (2019). Corpora in legal translation: Overcoming terminological and phraseological asymmetries between Czech and English, *CLINA 5*–2, 165–186.

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

L'Homme, M.-C. & Azoulay, D. (2020). Collecting collocations from general and specialised corpora: A comparative analysis. In G. Corpas Pastor & J.-P. Colson (Eds.), *Computational Phraseology* (pp. 151–175). John Benjamins.

Landau, S. I. (2001). *Dictionaries: The art and craft of lexicography* (2nd edition). Cambridge University Press.

Law, J. (Ed.). (2022), *Oxford dictionary of law* (10th edition). Oxford University Press. `https://www.oxfordreference.com/display/10.1093/acref/9780192897497.001.0001/acref-9780192897497`. [Accessed January 15, 2024].

Lehecka, T. (2015). Collocation and colligation. In J.-O. Östman & J. Verschueren (Eds.), *Handbook of pragmatics* (pp. 1–20). John Benjamins.

Lewandowska-Tomaszczyk, B. (1990). Meaning, synonymy, and the dictionary. In J. Tomaszczyk & B. Lewandowska-Tomaszczyk (Eds.), *Meaning and lexicography* (pp. 181–208). John Benjamins.

Lyons, J. (1981). *Language and linguistics. An introduction*. Cambridge University Press.

Matilla, H. (2006). *Comparative legal linguistics*. Ashgate.

Matulewska, A. (2016). Semantic relations between legal terms. A case study of the intralingual relations of synonymy. *Studies in Logic, Grammar and Rhetoric*, *45*(1), 16–174.

Mel'čuk, I. (1998). Collocations and lexical functions. In A. P. Cowie (Ed.), *Phraseology, theory, analysis and application* (pp. 23–53). Clarendon Press.

Michta, T. (2022). You shall know a term by the company it keeps: Collocations of the term evidence in general and legal corpora. *Beyond Philology: An International Journal of Linguistics, Literary Studies and English Language Teaching*, *19*/1, 65–96.

Michta, T., Kloza M., Łompieś J., Mela M., Mela W., Miąc M., Newska J & L. Religa (2009). Studencki słownik kolokacji angielskiego języka medycyny. In M. Łukasik (Ed.), *Debiuty Naukowe III* (pp. 89–225). Katedra Języków Specjalistycznych UW.

Michta, T. & Mroczyńska, K. (2022). *Towards a dictionary of legal English collocations*. Wydawnictwo Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach.

Money, J. (1955). Hermaphroditism, gender and precocity in hyperadrenocorticism: Psychologic findings. *Bulletin of the Johns Hopkins Hospital*, *96*, 253–264.

Nadler, J. & Mueller, P. A. (2017). Social psychology and the law. In F. Parisi (Ed.), *Oxford handbook of law and economics Vol.1. Methodology and concepts* (pp. 124–160). Oxford University Press.

Rzepkowska, A. (2023). The collocational profile of employment and work in UK employment law. *Conversatoria Linguistica XV*, 67–87.

Schiappa, E. (2022). *Transgender exigency. Defining sex and gender in the 21st century*. Routledge.

Siepmann, D. (2005). Collocation, colligation and encoding dictionaries. Part I: Lexicological aspects*. International Journal of Lexicography*, *18*(4), 409–443.

Siepmann, D. (2006). Collocation, colligation and encoding dictionaries. Part II: Lexicographic aspects. *International Journal of Lexicography, 19*(1), 1–39.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press.

Sinclair, J. (2004). *Trust the text. Language, corpus and discourse*. Routledge.

Stoller, R. J. (1964). A contribution to the study of gender identity. *International Journal of Psycho-Analysis*, *45*, 220–226.

Stoller, R. J. (1968). *Sex and gender: On the development of masculinity and femininity*. London: Hogarth Press.

Taylor, C. (2021). Investigating gendered language through collocation. In J. Angouri & J. Baxter (Eds.), *The Routledge handbook of language, gender, and sexuality* (pp. 572–586). Routledge.

Whorf, B. L. (1956). *Language, thought, and reality: selected writings*. Technology Press of Massachusetts Institute of Technology.

Wiegman, R. (2002). Academic feminism against itself. *NWSA Journal 14*(2), 18–37. http://www.jstor.org/stable/4316890. [Accessed January 6, 2024].

Więcławska, E. (2023a). Approaching legal multinomials from the sociolinguistic perspective–insights into authorship-based distinctions. *International Journal for the Semiotics of Law-Revue internationale de Sémiotique juridique, 36*.4, 1699–1715.

Więcławska, E. (2023b). *Binomials in English/Polish company registration discourse: The study of linguistic profile and translation patterns*. V&R Unipress.

Woodley, M. (Ed.). (2013). *Osborn's concise law dictionary*. Sweet & Maxwell Ltd.

Wouters, J. (2020). From an economic community to a union of values: The emergence of the EU's commitment to human rights. In J. Wouters et al. (Eds), The European Union and human rights: Law and policy (Oxford, 2020; online edition, Oxford Academic, 18 Feb. 2021). https://doi.org/10.1093/oso/9780198814191.003.0002 [Accessed February 21, 2024].

Yevchuk, A. (2021). Plesionyms as a vocabulary teaching tool: The case of Estonian EFL learners. *Sustainable Multilingualism*, *19*(1), p. 20–226.

Zlatev, J. & Blomberg, J. (2015). Language may indeed influence thought. *Frontiers in Psychology*, *6*, 1631. doi: 10.3389/fpsyg.2015.01631

Online sources

Cambridge Dictionary online. 2024. https://dictionary.cambridge.org/dictionary/english [Accessed January 6, 2024].

Collins English Dictionary online. 2024. https://www.collinsdictionary.com/dictionary/english [Accessed January 6, 2024].

Council of Europe: European Court of Human Rights, Handbook on European non-discrimination law. 2018. ISBN 978-92-871-9851-8 available at https://rm.coe.int/fra-2018-handbook-non-discrimination-law-2018-en/1680a2b52b [Accessed February 1, 2024].

Equality Glossary. Council of Europe. March 2016. https://edoc.coe.int/en/gender-equality/6947-gender-equality-glossary.html [Accessed January 6, 2024].

European Institute for Gender Equality, What is gender mainstreaming, https://eige.europa.eu/gender-mainstreaming/what-is-gender-mainstreaming?language_content_entity=en [Accessed February 6, 2024].

Gender Equality Glossary: Terms and Concepts. UNICEF. 2017. https://www.unicef.org/rosa/media/1761/file/Genderglossarytermsandconcepts.pdf [Accessed January 6, 2024].

The Law Dictionary: https://thelawdictionary.org [Accessed January 6, 2024].

https://www.europarl.europa.eu/factsheets/en/sheet/59/equality-between-men-and-women

WHO: https://www.who.int/health-topics/gender#tab=tab_1 [Accessed January 16, 2024].

Merriam-Webster.com Dictionary. 2024. https://www.merriam-webster.com/dictionary [Accessed January 6, 2024].

Siekiera, J. 2022. Gender equality in the European Union law – the use of the words “sex” and “gender” in the primary and secondary legislation. Available at https://efektywne-prawo.org.pl/gender-equality-in-the-european-union-law-the-use-of-the-words-sex-and-gender-in-the-primary-and-secondary-legislation-dr-joanna-siekiera/ [Accessed February 1, 2024].

Wex. Legal Information Institute Cornell Law School. https://www.law.cornell.edu/wex [Accessed January 6, 2024].

***

Katarzyna Mroczyńska is a linguist with a background in economics, with her main research interest being collocations and the legal English genre. Other academic interests include translation of specialist texts, recent developments in computer-assisted translation and machine translation, borrowings in Polish specialist language, the role of metaphors in Business English, neurolinguistics, and the application of cognitive field theory in foreign language teaching and learning.

AGNIESZKA RZEPKOWSKA[1]
University of Siedlce, Poland
https://orcid.org/0000-0003-3381-0870

DOI: 10.15290/CR.2024.45.2.06

Employee, *worker*, *jobholder*, *agent*, *staff* and *workforce* in UK employment legislation: A genre-specific corpus study on synonymy, collocations and meaning

Abstract. In legal texts, synonymy may lead to confusion, especially if the synonymous words are terms which, by definition, should be unambiguous. This paper addresses the issue of synonyms in legal language through a genre-specific corpus study of *employee*, *worker*, *jobholder, agent*, *staff* and *workforce* – legal terms that appear similar in meaning – in the corpus of UK employment legislation. Specifically, the study looks at (a) the distribution of the terms in the corpus to determine the areas of law in which they are used, (b) the definitions of these terms in legal dictionaries, as well as general and business English dictionaries if the legal dictionaries fail to provide definitions, along with legal definitions from the 12 legislative documents constituting the corpus, (c) the immediate context of use (the co-text) to identify the most typical word combinations with the terms (candidate collocates), and (d) the differences between the terms based on the definitions and the collocational profile[2] of the terms. The findings suggest that, to some extent, the meanings of the terms overlap, indicating that they function as synonyms. However, they are not interchangeable in legislative acts as indicated by both their distribution in the corpus and their immediate context. Additionally, the study identified not only candidate collocations but also several multi-word terms defined within the legal acts.

Keywords: collocation, corpus, legal discourse, legal language, multi-word term, synonym, word combination

1 Address for correspondence: Instytut Językoznawstwa i Literaturoznawstwa, Uniwersytet w Siedlcach, ul. Żytnia 39, 08-110 Siedlce, Poland. E-mail: agnieszka.rzepkowska@uws.edu.pl

2 A collocational profile is understood here as the immediate linguistic context (or co-text) in which the words tend to appear in the corpus. This profile includes the most frequent word combinations of the words subject to rules determined further in the paper (“The study”).

1. Introduction

Synonymy is a phenomenon generally not welcome in legal discourse, particularly where legislation is concerned. Lawyers typically associate each term with a distinct meaning and tend to avoid using synonyms. However, despite being frowned upon, synonyms, or near-synonyms, have nonetheless found their way into legal terminology and settled quite comfortably in legal texts (Klabal, 2022; Matulewska, 2016; Goźdź-Roszkowski, 2013).

From a pragmatic point of view, synonymy can be seen as a relation between two words in use – words that "map to the same meaning or concept" (Murphy, 2003, p. 145). Therefore, studying the context in which they appear may shed light on the differences between their meanings – sometimes subtle but still present and recognised by experts in a given field. The differences may not be obvious to non-experts for whom a dictionary definition, if available, may not be enough to illustrate the discrepancies in meaning. However, what can help illuminate these distinctions is seeing the words in their 'natural environment' – the texts in which they were originally used. In legal contexts, such texts are often legislative documents.

One avenue which can be taken to investigate the synonymous relations between words is to examine the other words that cooccur with the lexical units being studied. The nature of these cooccurrences may vary, depending on the strength of the connection between words and their mutual expectancy. A text can reveal a spectrum of different semantic relations between cooccurring lexical items, some loose, others more fixed – ranging from free combinations to collocations and ending with idioms (the most fixed of all three). Collocations, in particular, can provide valuable insights into the actual usage of words in both general and special-purpose languages. The fact that a word or a word combination collocates with a certain lexical item and not with others can illustrate part of its meaning. Legal language is no exception to this phenomenon.

Corpus tools are useful for identifying word combinations. They provide simple calculations of word frequencies and more complex statistical measures that make use of frequency counts. Some of the measures, such as MI-score, MI2-score, Delta P, Cohen's d and logDice – the latter used by default in Sketch Engine (Brezina, 2018, p. 70) – calculate the association measure between words in a corpus and are thus used to identify candidate collocations. Corpus analytics can also be used to compare words based on the lexical items they tend to co-occur with, helping to distinguish the use of synonyms in specific types of texts.

The paper explores whether genre-specific corpus studies could help answer questions about the use of near-synonymous legal terms in legal contexts using the example of *employee*, *worker*, *jobholder, agent*, *staff* and *workforce*. These six terms are semantically related and may be treated as synonyms or near-synonymous in the context of UK employment law. Specifically, the study investigates:

— whether genre-specific corpus studies can provide insights into the specific area of law in which each of the terms is used;

— whether the terms are synonyms and, if yes, how closely related they are; and
— whether a list of the most frequent word combinations with the terms (candidate collocates) compiled using corpus tools can reveal, or contribute to revealing differences between them.

Following a literature review on synonymy in legal language, legal discourse and collocations, this paper proceeds with the four steps listed below:

1. examining the distribution of the terms in the corpus to identify the areas of employment law where they are most frequently used,
2. finding and comparing dictionary definitions of the terms with their legal meanings as established in UK employment legislation;
3. analysing the immediate context in which the terms appear in the corpus by identifying the strongest word combinations (with candidate collocates), using the logDice score calculated by Sketch Engine; and
4. differentiating between candidate legal collocations and multi-word terms, and interpreting the meaning of the terms based on the word combinations found in the corpus.

2. Legal language and legislative documents

There is an inextricable connection between law and language. Everything that law encompasses needs to be expressed in language and vice versa: language is the primary means of understanding law. As law applies to and is applied in various walks of life, its language naturally varies in different contexts. Goźdź-Roszkowski aptly observes that "what is routinely referred to as 'legal language', represents an extremely complex type of discourse embedded in the highly varied institutional space of different legal systems and cultures" (Goźdź-Roszkowski, 2012, p. 1). He also elaborates on a range of categories of legal language discussed in the literature, including:

— frozen, formal, consultative and casual types of written texts (Danet, 1980, p. 371);
— judicial discourse, courtroom discourse, language of legal documents and the discourse of legal consultation (Maley, 1994, p. 16);
— the language of law (the language of legislation and contracts) and other uses of "legal language" (Trosborg, 1995, p. 31).

Similarly to Trosborg, Wróblewski distinguishes between two types of legal language: the language of the sources of law, that is legislature and contracts (in Polish *język prawny*), and the metalanguage of law, the language used, among others, by lawyers to talk about law (in Polish *język prawniczy*) (Wróblewski, 1948). Bhatia (2006) differentiates between the primary legal genre, which includes primary sources of law, and the secondary legal genre, which comprises a reproduction of legislation that exhibits a high degree of intertextuality with the primary sources[3]. De Groot proposes a similar division, focusing not on legal language in general but on legal vocabulary: (i) the vocabulary employed by legislators in acts of law, (ii) the vocabulary used by

3 Klabal (2019, p. 167) relates to them as higher-order genre and lower-order genre, respectively.

lawyers of a legal system and in commentaries on that legal system, and (iii) terminology found in general publications concerning that legal system (De Groot, 1996, p. 378). In light of this, we can classify legislative documents as written texts that use the language of law as defined by Trosborg and Wróblewski, featuring vocabulary specifically used by legislators.

Legal language is noted for its precision, indeterminacy, specialization, complexity and conservatism (Goźdź-Roszkowski, 2012). These features characterise both the primary and secondary legal genres. Precision is exemplified, among others, by the use of legal terms: words or phrases of specific meaning outlined in a given legal act. Indeterminacy allows lawyers to adjust the interpretation of a legal act to current circumstances. Specialization indicates that a text employs a language designed for special purposes. Complexity pertains, among others, to the intertextuality of legal texts in the continental systems of law or the all-inclusiveness of documents in the common law system. In English legal language, conservatism is manifested through traditional grammar, which includes old-fashioned words and phrases, such as modals and semi modals like *shall* and *may*, formal adverbs such as *whereby*, *therein*, *hereby*, etc., and the use of historically-formed expressions like doublets and triplets.

All those features make legal discourse very unique. On the one hand, precision is required, on the other, indeterminacy is also present. In the context of this paper, these two features deserve special attention. Seen, among other things, in a consistent use of legal terminology, precision appears to exclude the possibility of using absolute synonyms in a legal text, particularly in legislative documents. On the other hand, it is this same precision that paves the way for near-synonyms, or terms that semantically seem very close but differ contextually. The context here may refer to a given branch of law, legal genre, particular enactment, its section, paragraph or a single sentence. Indeterminacy may be employed by the legislator on purpose to allow a great deal of freedom in interpreting certain provisions when generalisation and flexibility of solutions are needed, and when neither precision nor determinacy is necessary or recommended (Kaczmarek, 2013, pp. 55–56).

There is no discussion about legal language without mentioning its core element: legal terms. In general, a term is a word or a word combination that represents a specific concept in a specific terminological system (Lukszyn, 2001, pp. 9–14). The main features of terms include:

- specialisation, which is evident in its use by specific users in specific situations in reference to specific objects;
- conventionality, which arises from the fact that terms are not formed in the course of a natural linguistic process but are the outcome of a purposeful action by a specific group of professionals;
- system-based meaning, which means that each term is part of a specific terminological system;
- accuracy and explicitness, which means that each term is defined in a specific way within a given professional context; and
- neutrality in terms of emotive and stylistic features (Kornacka, 2005, p. 131).

The status of a term in law-making acts comes from those documents where it is either directly or indirectly defined. The context of the legislative document can help to reconstruct the meaning, or the co-text of the term may indicate that the word or phrase has a specific system-based meaning (Rzepkowska, 2021, pp. 20–21). In addition to collocations or free combinations, multi-word terms are word combinations used in legal documents and their use is rarely incidental. However, distinguishing between a term and a non-term can be problematic as the difference between a term and a phraseme in legal texts may become obscure (Biel, 2012, p. 227) and requires an in-depth systemic and contextual analysis taking into account the features listed above.

An analysis of word combinations comprising terms must consider the phenomenon of nested terms. These are one-word or two-word terms that form part of longer terms (Marciniak & Mykowiecka, 2014a; Marciniak & Mykowiecka, 2014b) like in *fixed-term employee*, where both *employee* and *fixed-term employee* are legal terms in the relevant statutes[4]. Therefore, it is important to bear in mind that a term found in a corpus may also be part of larger multi-word terms, thus going beyond a typical collocation. It is possible to initially verify with special algorithms whether a given word combination is a term. For instance, when discussing nested terms, Marciniak and Mykowiecka suggest taking into account the frequency of occurrence of terms in isolation and nested, as well as the number of different contexts in which the whole nested phrase appears (Marciniak & Mykowiecka, 2014b, pp. 2–3). However, the final decision should be based on the context of use. In legal texts, the context often provides a definition of a term, offering a definite proof that a given phrase is a term. However, often does not mean always. The issue becomes more complicated when a legal term is not defined in a given statute or a number of statutes regulating a specific matter, but is still regarded by lawyers as a legal term. As Rycak observes in relation to Polish law on working time regulations, there are no definitions of a number of terms that the legislator uses when regulating working time. Additionally, the legislator uses certain terms illogically. All of that leads to numerous disputes on fundamental issues in legal academic writings (Rycak, 2008, p. 15). In such cases, determining whether a lexical item is a term requires a thorough analysis of its usage context, supported by in-depth, expert knowledge of the subject matter.

3. Synonymy

Synonymy has a long history of research in linguistics and has been characterised in various ways. Lyons distinguishes between complete synonymy and absolute synonymy. The former is defined as two items sharing descriptive, expressive and social meaning, and the latter as two items that feature "the same distribution and [being] completely synonymous in all their meanings and in all their contexts of occurrence" (Lyons, 1981, p. 148). While complete synonymy is rare, absolute synonymy, as defined in this way, is considered nearly non-existent in natural language

4 The example of *fixed-term employee* has been taken from the corpus under analysis.

as words are rarely fully interchangeable in all contexts. Clark argues that language eliminates absolute synonyms because, over time, one word in a synonymous pair tends to fall out of use or take on a slightly new meaning (Clark, 1992, p. 177; see also Ullmann, 1972, p. 141). Scholars also note a scale of synonymy, with some words being very close in meaning to one another and others being more distant (Cruse, 1986, pp. 265–266; Goźdź-Roszkowski, 2013, pp. 96–97).

English legal language provides us with a specific type of synonymy, which manifests in the phenomenon of doublets, triplets and even quadruplets (Crystal, 2006) in legal texts[5]. These are expressions representing very similar or complementary concepts, belonging to the same grammatical category and usually joined by *and* or *or* (based on Carvalho, 2008, p. 1; Bhatia, 1993, p. 108). Vázquez y del Árbol (2006, 2014) notes that they tend to appear particularly in notarial documents and testaments, though they are present in other types of texts as well, for instance contracts (Carvalho, 2008). Discussing doublets, Buşilă observes that "[e]ven though these constructions are etymologically validated, semantically – they become tautologies or pleonastic phrases: *ideas and opinions*, *null and void*, *defamatory or untrue*, *relevant and sufficient*, *unreasonably or arbitrarily*, *final and unappealable*, *costs and expenses*, etc." (Buşilă, 2016, p. 190) and they are used "more as an incantation than for any legal reason" (Buşilă, 2016, p. 190). Adams shares this view, adding that the use of certain strings of synonyms or near-synonyms like *sell, convey, assign, transfer, and deliver* in a purchase agreement may be an example of improvisation and result from "[finessing] the often-awkward task of selecting the best word for a given provision" (Adams, 2004, p. 204).

This phenomenon of "redundant synonyms" (Adams, 2004, p. 204) may serve as an example of synonymy rooted in legal practice and contrasts with the general opinion that synonymy in its extreme form is generally undesirable in legal practice. For example, Jopek-Bosiacka observes that:

> [i]n order for the law to function, the principle of semantic accuracy or language consistency must be observed. Once a technical term was selected, it must be repeated over and over again instead of using synonyms. The use of synonyms is discouraged in legal texts because the user might think that reference is being made to a different concept. (Jopek-Bosiacka, 2011, p. 16)

However, a number of studies, often carried out with corpus-analysis tools, show that synonymy of varying degree exists in legal contexts (Goźdź-Roszkowski, 2013; Matulewska, 2016; Cao, 2007). Researchers' interest in legal synonyms spans different areas of law, such as contract law (Biel, 2012) and competition law (Biel & Koźbiał, 2020); different legal systems, such as the American legal system (Goźdź-Roszkowski, 2013) and the EU legal system (Biel & Koźbiał, 2020); different areas in which synonyms are an issue, such as translation (Cao, 2007; Chroma, 2011; Jopek-Bosiacka, 2011) and intralingual synonymy (Matulewska, 2016); and different legal texts,

5 Bhatia (1993) refers to them as binomial and multinomial expressions.

such as notary acts (Vázquez y del Árbol, 2014; Buşilă, 2016) or agreements (Calvalho, 2008). The corpora used by researchers often consist of certain types of texts, such as legislative documents, or a collection of texts of different legal genres.

The phenomenon of synonymy in legal texts is usually defined in a very broad sense. Discussing legal language, Matilla explains that synonymy is when "two or several terms express the same concept" (Matilla, 2006, p. 144). In her study on the semantic relations between legal terms, Matulewska defines synonymy as "a semantic relation that binds two terms with the same referential meaning – but not necessarily the same pragmatic meaning – which belong to the same part of speech and differ in spelling" (2016, p. 163).

As synonymy can be viewed across a broad spectrum, there are numerous ways of classifying it in the literature. Matulewska distinguishes between interlingual synonyms, corresponding to equivalence as addressed in translation studies, and intralingual synonyms, which are words with similar meanings in the same language. She also lists different situations in which synonymous terms occur: (1) between different languages: (a) in vernacular and legal languages, (b) in legal and other special purpose languages; and (2) within legal language: (a) due to a lack of terminological consistency, (b) due to the passage of time (synonymous terms in diachronic perspective), (c) due to different text genre (legal-genre dependent synonymy), and (d) due to different branches of law (branch-of-law-dependent synonymy) (for examples of each type of synonyms see Matulewska, 2016, pp. 164–170). Basing her classification on Murphy (2003), Chromá discriminates between lexical synonymy (relations between lexical units) and propositional synonymy (relations between syntactic units) also referred to as paraphrase (Chroma, 2011, pp. 39–40). Klabal makes a distinction between synonyms across parts of speech and categorises synonyms into: adjectival synonyms, verbal synonyms, nominal synonyms and prepositional synonyms (Klabal, 2022, pp. 72–73).

A concept of synonymous relations often referred to by scholars is plesionymy. The term relates to situations where words are not exchangeable in all contexts; they are very similar but not identical in meaning, with their denotation, connotation, implicature, emphasis and register possibly varying (Edmonds & Hirsty, 2002, p. 107; Yevchuk, 2021, p. 204; Austin, 1962). Cruse (1986, pp. 285–286) categorizes plesionyms alongside absolute synonyms and cognitive synonyms, noting that among these three synonym categories, plesionyms are the least similar on the scale of similarity. In fact, plesionymy is useful for capturing various relations, like hypernymy and hyponymy, between different legal terms and ordinary words. Jopek-Bosiacka, writing about legal terms that have acquired their general meaning from ordinary words, gives such examples as *assault*, *battery*, *purchase*, or *domicile* (Jopek-Bosiacka, 2011, p. 11), showing that the way they are understood in legal language differs from their general use. Yet, they do remain synonymous to some extent, and thus can be treated as examples of hypernyms and hyponyms.

Legal synonyms are a subject of research in translation practice. Cao provides a number of examples of legal synonyms, and points out that they are not identical because what makes them different is their connotations (e.g. in the case of *encumbrance*, *mortgage*, *charge*, *pledge*,

lien), the type of legal writing they are used in and the area of law they pertain to (2007, pp. 71–73). In discussing legal translation practice, she illustrates that identifying and ascertaining the legal meaning of a word in relation to its general meaning takes place through an analysis of the context the word occurs in. "This includes both the wider legal context, such as a particular area of law, and the immediate linguistic context such as a sentence, the paragraph and the entire text in which the word is used" (Cao, 2007, p. 70). Examining the context is particularly important as there are terms in legal texts that originate from general language "but are assigned a special legal meaning by each legal system" (Jopek-Bosiacka, 2011, p. 10). Legal synonymy in the context of translation is also studied by Biel and Koźbiał, who explore near-synonymous legal terms in EU English-Polish competition law (Biel & Koźbiał, 2020). They investigate source-text synonymy in translation and conclude that "source-text synonymy causes variation and that, due to asymmetries between languages, it is difficult to control synonymy and standardise variants in translation" (Biel & Koźbiał, 2020, p. 87).

This short literature review brings us to a conclusion that synonymy is a gradable phenomenon, where instances of absolute synonymy are rare, particularly in legal language. More frequent are near-synonyms or plesionyms (as some scholars call them). Also, learning about synonymy entails investigating the context in which two or more lexical units appear. Lastly, despite being omnipresent, synonymy, particularly in its extreme form, is generally viewed as an undesirable phenomenon in legal discourse, and rarely allows for full interchangeability.

4. Collocations

So natural to native speakers and at the same time so hard to acquire in foreign-language learning, collocations are an essential component of every language. The term collocation is credited to Firth (1957) who first used it in his linguistic theory. Burkhanov states that the concept is used to "account for characteristic word combinations which have acquired an idiomatic, or rather semi-idiomatic, semantic relationship because of their frequent cooccurrence in the *context* (1), as *dog* and *bark*, *dark* and *night"* (Burkhanov, 1998, p. 39). Cruse, on the other hand, refers to collocations as habitually cooccurring "sequencies of lexical items" (Cruse, 1986, p. 40), in which each constituting lexeme is semantically transparent as opposed to an idiom, the meaning of which is not a direct derivative of its constituents. The *Dictionary of Lexicography* defines collocation as "the semantic compatibility of grammatically adjacent words" (Hartmann & James, 1998, pp. 22–23). This definition pays attention to the patterns of cooccurrence of words such as adjective-noun, noun-verb and verb-preposition. Hartmann and James suggest that collocations should be viewed in opposition to idioms and free combinations as they are less fixed than the former and more fixed than the latter. Cowie takes a similar approach placing collocations next to idioms and quasi-idioms under the category of semantic phrasemes, treating them as the least fixed of the three types (Cowie, 1998, p. 30). The dividing lines between free combinations and collocations, and between collocations and idioms are sometimes blurry, with many phrases existing on the borderline between these categories.

Frequency as a means of assessing collocations was anticipated by Firth (1957/1968, p. 180) and Halliday (1961, p. 277). The increasing accessibility of corpus analytics offering various quantitative formulas in which frequency is the core element has recently made the frequency-based approach a feasible tool for finding and evaluating collocations. The calculation of the probability of cooccurrence and the resulting association measure is done with the use of such data as the number of tokens in the corpus, the frequency of the node, the collocate and the collocation as a whole, i.e. the node and collocate, in the whole corpus and the collocation window size (Brezina, 2018, p. 70). On the other hand, the phraseological approach views collocation as an association of lexemes that tend to occur in certain grammatical constructions. The meaning of such word combinations should be transparent (derivable from the meanings of the constituting lexemes), which distinguishes them from idioms. They are lexically variable but the selection of individual components, that is the collocates of the node, can be restricted at one or more points (Cowie, 1994, pp. 115–116; Sinclair, 2005), which makes them also different from free combinations. Mitigating the dispute concerning which approach to collocations is best, Michta and Mroczyńska (2022, p. 14) justly note that each of the approaches answers different questions and thus they should not be perceived as two opposing methods but as partners walking hand in hand. Hence neither should be called "empty" (Siepmann, 2005, p. 411) because if one of them is, the other may be as well.

5. The study

To shed some light on the information that can be reconstructed from the collocational profile of legal terms, this analysis will focus on six words found in UK employment legislation: *employee*, *worker*, *jobholder, agent*, *staff* and *workforce*. These terms will be examined alongside their synonymous relationships based on definitions from both dictionaries and the legal contexts in which they appear. As the selected lexical units are nouns, they will be studied from the viewpoint of lexical synonymy, particularly nominal synonymy. Nominal synonyms in legal language are "often terminological in nature and usually cannot be used interchangeably, or may also be a case of collocations or jurisdictional variation" (Klabal, 2022, p. 72). Therefore, I do not expect to find examples of absolute synonymy (Lyons, 1981) in this analysis. The question is whether there are instances of complete synonymy or near-synonymy among these terms.

Considering the type of texts in the corpus, comprising 12 UK statutes on various employment-related issues, the analysis is also expected to yield findings on legal-genre-dependent synonymy. That is in line with Lyons' assertion that "frequent use of a word or phrase in one range of contexts rather than another tends to create a set of associations between that word or phrase and whatever is distinctive about its typical contexts of cooccurrence" (Lyons, 1981, p. 150). In this research, the context is understood in two ways: as the whole text in which the terms are found, and the very close environment in which the terms appear and where their collocates are found – also referred to by scholars as co-text (Halliday, 1999, p. 3). A review of word combinations – ranging from free combinations to relatively fixed specialised collocations

and multi-word terms – serves as a means to understand their typical context as employed by the legislator.

5.1. Data and methods

5.1.1. The corpus

This paper uses the term "corpus" in the sense proposed by Sinclair as "a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" (Sinclair, 2005, p. 16). The corpus tool selected for the analysis is Sketch Engine[6] which processes texts to present results in the form of Word Sketches, concordances and word lists, among others. This research takes advantage of all these functions to learn about the collocations of words under analysis and unlocks the potential of context as a medium for filling gaps in understanding differences between near-synonyms.

The study relies on the author's corpus of legal texts compiled according to the predefined criteria, including the legal system, legal domain and type of texts. The corpus comprises legislative documents governing employment relations in the United Kingdom. These documents were selected based on the UK government websites[7] and commercial legal websites[8] offering information and advice on employment regulations in the UK. The laws referred to by experts as the most important where subsequently found on the official legislation website, which includes all enacted legislation for the UK, Scotland, Wales and Northern Ireland[9]. The final corpus consists of 12 documents, 1.2 million tokens[10] and over 760 thousand words[11] (Table 1).

Table 1. UK Employment Law corpus composition

Legislative document	Tokens	Percentage of the corpus
Trade Union and Labour Relations (Consolidation) Act 1992	303,082	25%
Employment Rights Act 1996	278,938	23%

6 See https://www.sketchengine.eu.

7 See https://www.gov.uk/browse/employing-people/contracts.

8 See https://croner.co.uk/resources/employment-law/legislation-list/ and https://www.expatica.com/uk/working/employment-law/employment-law-uk-104502/.

9 See https://www.legislation.gov.uk.

10 Sketch Engine defines a token as the smallest unit that a corpus consists of. A token may refer to: a word form, punctuation, a digit, abbreviations, and anything else between spaces (https://www.sketchengine.eu/my_keywords/token/).

11 Sketch Engine defines a word as a type of token which begins with a letter of the alphabet (https://www.sketchengine.eu/my_keywords/word/).

Legislative document	Tokens	Percentage of the corpus
Equality Act 2010	203,351	17%
Health and Safety at Work etc. Act 1974	160,553	13%
Pensions Act 2008	114,655	9%
National Minimum Wage Act 1998	49429	4%
Working time regulations 1998	40,501	3%
Agency Workers Regulations 2010	22,465	2%
Transfer of Undertakings (Protection of Employment) Regulations 2006	14,796	1%
The Fixed-term Employees (Prevention of Less Favourable Treatment) Regulations 2002	9,772	1%
The Maternity and Parental Leave etc. Regulations 1999	9,175	1%
Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000	6,959	1%
Total	**1,213,676**	**100%**

Table 1 lists legislative documents under analysis dealing with the issue of employment relations in the UK, arranged according to their size. It is worth noting that the two top documents are particularly large, making up nearly half of the corpus (48%). They are followed by three that constitute another 39%. The remaining 13% comprises seven acts that are much shorter, ranging from 4% to 1% of the corpus.

5.1.2. Selection of words for analysis

The study covers the legal term *employee* and its synonyms in legislative documents governing employment relations in the United Kingdom. This term was selected for analysis because it is one of the fundamental terms underlying employment relations in the UK legal system. A simple search using the Sketch Engine Wordlist tool, which excludes non-words, revealed that *employee* ranks 23rd in a list of nouns, positioned behind *employment* (10th) and *employer* (19th). Other nouns that ranked higher than *employee* were mainly words responsible for text organisation, such as *paragraph*, *section*, etc.

A review of the corpus made it possible to single out five synonyms of the term *employee*. The search for synonyms relied on a list of synonyms found in online thesauruses: thesaurus.com[12],

12 employee. (n.d.) *thesaurus.com*. https://www.thesaurus.com/browse/employee.

Cambridge thesaurus[13] and Collins thesaurus[14]. All nouns and noun phrases listed there as synonyms were searched for in the corpus. Synonyms with the absolute frequency in the whole corpus exceeding 50 were subject to further analysis. Only five words met these criteria: *worker*, *jobholder*, *agent*, *staff* and *workforce*.

The classification proposed by Benson et al. (2009, p. XXXI-XXXIV) has served as an inspiration for classifying the collocations found in the corpus. However, it was modified to meet the specific objectives of this study by including the relation of possession. The discussed classification presents different configurations of the node and collocate; the node is one of the terms under analysis: *employee*, *worker*, *jobholder*, *agent*, *staff* and *workforce*. The collocate can be a verb, adjective or noun (a premodifier), or a different noun (a possessive relation). The following groups of word combinations are further analysed:

— premodifier + NODE,
— NODE + noun,
— verb + NODE (object),
— NODE (subject) + verb, and
— NODE's+ noun.

Since the selected terms are nominal synonyms and as such may be terminological in nature and not interchangeable, it is assumed that the selected words are not freely used by the legislator, but their choice is dictated by their specialist meaning. Additionally, they are not examples of jurisdictional variation as all the texts in the corpus come from the UK legal system.

5.2. Dictionary and legal definitions of terms under analysis

This part of the paper covers an analysis of dictionary and legal definitions of the words. Legal definitions were looked for in two legal dictionaries: *The Penguin Dictionary of Law* (Webb, 2010) and *A Dictionary of Law* (Martin, 2003). Words which were not found in those dictionaries were looked up in an online general English dictionary and a business English dictionary: *Collins English Dictionary*[15] and *Cambridge Business English Dictionary*[16].

A review of selected dictionaries showed that three out of six words under analysis have an entry in legal dictionaries, which proves their status as well-established legal terms. These are *employee*, *worker* and *agent*. The remaining three, *jobholder*, *staff* and *workforce*, seem not to be terminographically-recognised legal terms as they are found in non-legal dictionaries only.

A review of the statutes (the corpus) shows that not only are *employee*, *worker*, and *agent* legal terms, but so are the other three terms under analysis because they have been defined by

13 employee. (n.d.) dictionary.cambridge.org/thesaurus/. https://dictionary.cambridge.org/thesaurus/employee.

14 employee. (n.d.) *Collinsdictionary.com*. https://www.collinsdictionary.com/dictionary/english-thesaurus/employee.

15 Available at https://www.collinsdictionary.com/.

16 Available at https://dictionary.cambridge.org/.

the legislator, either directly or indirectly. What should be stressed is that such legal definitions are usually applicable only within the act of law in which they appear unless otherwise stated, or in another act when there is a direct reference to that other document. It is common to find phrases such as *in these regulations* appearing next to the definitions, and phrases like *has the meaning assigned by section*, *within the meaning of* or *has the meaning given in*[17] next to terms, which clearly delimits the use of the definitions and clarifies the meaning of the terms.

Employee is a word encountered in both legal dictionaries reviewed. The term is defined as a person working under the direction or control of another for a wage, salary, fee or other payment. The relationship between the employee and the employer is governed by an employment contract. Employees enjoy special rights and are protected by certain laws that no contract of employment may deprive them of (this is the most protected form of employment under employment law). This definition of *employee* seems to be based on the legal definitions found in the statutes (for instance in the Employment Rights Act 1996).

Worker is usually defined in legal dictionaries as a person employed to do work for another under an employment contract or any other contract. On the other hand, *A Dictionary of Law* (Martin, 2003) does not provide a definition of *worker* but equates the word directly with *employee*, thus making *worker* a close synonym of *employee* in legal terms. Legal sources are more specific. For example, the Employment Rights Act 1996 states that the worker performs personally work or service, determines the form of the contract (oral or written) and differentiates a worker's contract from the relationship with a client or customer.

Agent is defined as a person appointed by another (the principal) to act on his/her behalf to perform a service, usually to negotiate a contract between the principal and a third party (Martin, 2003). Such a definition of an *agent* makes him/her a type of a representative. He or she is subordinate to the principal, the same as an employee to the employer, yet his or her liability and rights largely depend on the type of agreement with the principal, which makes an agent different from an employee. *Agent* is defined either directly or indirectly in a few statutes. In two cases, Employment Rights Act 1996 and Equality Act 2010, an agent is a type of a middleman, facilitating the conclusion of an agreement. On the other hand, an agent in the context of trade unions is understood as "a banker or solicitor of, or any person employed as an auditor by, the union or any branch or section of the union" (Section 119 of the Trade Union and Labour Relations (Consolidation) Act 1992). That meaning in fact mentions an employment relationship between an agent and trade unions, in which an agent can be an employee (an employed auditor).

Jobholder is not found in legal dictionaries. In general English dictionaries it is defined as someone who has a regular post of employment in an organisation. *Jobholder* is also a legal term of very narrow applicability (encountered in one statute, that is the Pension Act 2008). There it is precisely defined as a worker who is working or ordinarily works in the UK under the

17 The examples have been taken from the corpus under analysis.

worker's contract, who is aged at least 16 and under 75, and to whom qualifying earnings are payable by the employer in the relevant pay reference period.

Similarly *staff* has no entry in legal dictionaries. Its general English and business English meaning is all people working for a particular company or in a particular place. *Staff* is indirectly defined in a few statutes. The National Minimum Wage Act 1998 and Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000 provide us with definitions of *a relevant member of the House of Commons staff* and *a relevant member of the House of Lords staff*. The former means "any person who is employed under a worker's contract with the Corporate Officer of the House of Lords" (National Minimum Wage Act 1998), the latter "any person (a) who was appointed by the House of Commons Commission; or (b) who is a member of the Speaker's personal staff" (National Minimum Wage Act 1998, Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000). Based on the above we can conclude that in the studied corpus a member of staff is a person working in the UK Parliament, either employed or appointed. Therefore, there is a semantic relation between *a member of staff* and *an employee* relying on the subordinate nature of the two and the element of work performed by them.

Workforce is the third word that is not taken as a legal term by those compiling legal dictionaries. Both general and business English reference books define it in two ways: firstly, as the total number of people in a particular area, e.g. a country, who are available for work, and secondly, as the total number of people who are employed by a particular company or organisation or who are engaged in a specific activity. *Workforce* is indirectly defined in the corpus through the term *relevant members of the workforce* understood as "all of the workers employed by a particular employer, excluding any worker whose terms and conditions of employment are provided for, wholly or in part, in a collective agreement" (Working Time Regulations 2010). Further in the documents we find that *representatives of the workforce* "are workers duly elected to represent the relevant members of the workforce". Similarly to *jobholder*, *workforce* is a legal term of narrow applicability in UK employment law.

To sum up, the definitions reviewed indicate that the six terms seem to be synonyms with varying degrees of similarity, but they appear not to be interchangeable in the legal context. *Employee*, *worker*, *jobholder*, *agent*, *staff* and *workforce* share certain characteristics, among others the fact that they refer to individuals who perform work for a hiring person under an employment contract or other type of a contract (they are bound by contractual relations); they are also subordinate to someone. A special case exists for *staff* and *workforce*, which are collective nouns that refer to certain groups of workers or employees.

5.3. The six terms in the corpus

The distribution of the terms under analysis in each legislative document reflects the real use of the terms in the contexts of individual employment law acts (Table 2). *Employee* and *worker*

stand out in terms of frequency in the studied sample, amounting to nearly 2,000[18] per million tokens each, while the remaining four range from 175 in the case of *jobholder* to 47 in the case of *agent*.

Employee is a term preferred over *worker* in five acts, and *worker* is favoured over *employee* in seven acts. *Jobholder* stands out in this group as it is used only in one act from the corpus, the Pension Act 2008. *Workforce* is found in 6 acts out of 12. The words *agent* and *staff* are encountered nearly in the whole corpus, apart from the Maternity and Parental Leave etc. Regulations 1999 (where only *employee* and *workforce* are used).

Table 2. Distribution of *employee, worker, jobholder, agent, staff* and *workforce* and their frequencies in the corpus

Legislative document	employee	worker	jobholder	agent	staff	workforce
1. Trade Union and Labour Relations (Consolidation) Act 1992	371	510	0	7	43	0
2. Employment Rights Act 1996	1,355	785	0	12	32	7
3. Equality Act 2010	53	58	0	12	16	0
4. Health and Safety at Work etc. Act 1974	36	0	0	1	7	0
5.Pensions Act 2008	37	118	213	6	11	0
6. National Minimum Wage Act 1998	35	190	0	10	11	0
7. Working time regulations 1998	17	325	0	5	16	28
8. Agency Workers Regulations 2010	51	286	0	1	9	1
9. Transfer of Undertakings (Protection of Employment) Regulations 2006	169	0	0	1	2	7
10. The Fixed-term Employees (Prevention of Less Favourable Treatment) Regulations 2002	145	5	0	1	9	18
11. The Maternity and Parental Leave etc. Regulations 1999	114	0	0	0	0	21

18 All frequency numbers refer to singular and plural forms taken together.

Legislative document	employee	worker	jobholder	agent	staff	workforce
12. Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000	20	94	0	1	13	0
Absolute frequency	2,403	2,371	213	57	169	82
Average frequency per million tokens	1,980	1,954	175	47	139	68

5.3.1. Free combinations, specialised collocations and multi-word terms

This part of the paper presents the main word combinations with *employee*, *worker*, *jobholder*, *agent*, *staff* and *workforce* in the corpus, including free combinations, collocations and multi-word terms. Candidate collocations in the direct linguistic context are not verified in terms of their strength, and some may be considered to be merely free combinations. Yet, their existence in the corpus adds to the candidate collocational profile of the terms and may shed some light on their meaning. Therefore, they are referred to as collocations in this research. Main collocates are defined as those at the top of the lists in word sketches, sorted by logDice score[19] with an absolute frequency of at least 5 in the case of *employee* and *worker*. No frequency threshold is set for the remaining four nouns due to their relatively low frequency and a limited number of possible collocates.

Table 3 presents the collocates of *employee* and *worker* in five groups, based on the grammatical relationship between the node and collocate. The collocates in each group are listed according to the logDice score, starting with the highest. The selected collocates feature an absolute frequency of at least 5. Words that collocate with both *employee* and *worker* have been bolded.

Table 3. Collocates of employee and worker in the corpus (with an absolute frequency exceeding 5)

NODE	EMPLOYEE	WORKER
premodifier + NODE	affected, fixed-term, **comparable**, permanent, pre-TURERA, deceased, existing, **relevant**	young, part-time, opted-out, full--time, **comparable**, betting, protected, offshore, **relevant**, former, mobile, agency, shop, night, applicant, contract, zero hours, home

19 logDice is a statistic measure for identifying co-occurrence. Sketch Engine applies it to identifying collocations as its value indicates the typicality (or strength) of the collocation based on the frequency of the node and the collocate and the frequency of the whole collocation. Theoretically, the higher the score the more typical the collocation. The maximum logDice score value is 14. (https://www.sketchengine.eu/my_keywords/logdice/).

NODE	EMPLOYEE	WORKER
NODE + noun	representative, shareholder, share, information	
verb + NODE[20]	**entitle**, dismiss, select, **employ**, **permit**, **require**, concern, **allow**, **pay**, represent, engage, exclude, involve, qualify, reinstate, suspend, leave, **treat**, notify, re-engage, assign, lay off, affect	**employ**, **entitle**, **require**, **pay**, subject, **permit**, provide, propose, represent, involve, **treat**, **allow**, inform, give, affect
NODE + verb[21]	satisfy, **work**, intend, exercise, propose, die, hold, terminate, **refuse**, elect, sustain, present, return, follow, start, apply	**work**, constitute, fall within, carry out, signify, believe, receive, join, complete, remain, become, **refuse**
NODE'S + noun	**contract** (of employment), **employment**, period (of employment), entitlement, remuneration, right, application, death, notice	**contract**, wage, employer, **employment**, year, agreement

The group of collocations where the collocate modifies the node contains a large number of multi-word legal terms that are either directly or indirectly defined in the statutes and used as independent terms throughout individual documents: *fixed-term employee*, *permanent employee*, *pre-TURERA employee*; or in the case of worker: *agency worker*, *shop worker*, *young worker*, *part-time worker*, *opted-out worker*, *night worker*, *applicant worker*, *contract worker*, *zero-hours worker*, *offshore worker*, *mobile worker*, *home worker*. They are examples of terms nesting other terms; in this case the nested terms are *employee* and *worker*. The expressions *affected employee*, *comparable employee*, *deceased employee*, *existing employee*, and *comparable worker*, *protected worker* and *former worker* seem to be collocations (as legal definitions of these collocations have not been found in the corpus, it has been assumed that they should not be given the status of a term).

Both multi-word terms and specialist collocations are also found in the group where the node modifies a noun. The expression *employee shareholder* is a legal term, while the remaining three collocates form specialised collocations of varying strength.

Verb collocates serve as a source of information about the meaning of the nodes as the use of a given verb attributes certain characteristics to the node or indicates that certain characteristics have been attributed to the node by the legislator. An examination of the verbs with *employee* and *worker* as objects reveals that these terms should be at least to some extent synonymous, as both refer to individuals who are *employed* (which means they are involved

20 The verb *be* has been excluded from the list as being too common to form a typical collocation without an additional object that would make it specific.

21 The verbs be, *have, take, give, do* and *make* have been excluded from the list as being too common to form a typical collocation without an additional object that would make them specific.

in a contractual relationship in which they occupy a subordinate position); *entitled to*, *permitted*, and *allowed* something (which means have some rights); *paid* (which means they receive remuneration for what they provide); *required to* do something (which means that they do not act freely); and *affected* by something (which means that they are influenced by their broadly-meant environment). The information conveyed by these verbs aligns with the definitions of *employee* and *worker* presented earlier in this paper. Interestingly, an *employee* is much more frequently *dismissed* and *engaged* in something (based on the absolute frequency); and only an *employee* is *laid off*, *suspended*, *left*, *reinstated* (all the verbs relate to either employment or the post taken); and *assigned* something. Conversely, only a *worker* is *subjected to* something, *provided with* something, *proposed* something and *informed* about something. Additionally, the verb *to employ* is used twice as often with *worker* as with *employee*.

The group of collocations in which *employee* and *worker* are the subjects illustrates how these two types of individuals can act. Both of them can *work* and *refuse to* do something (for instance to work). However, the other verbs are typically associated with only one of the two. An *employee* has the power to *exercise* rights, *satisfy* needs, *intend*, *elect*, *apply*, *hold*, for instance a position, and *terminate* an agreement (in other words express his/her will). An *employee* needs to *return* something or somewhere and *follow* procedures.[22] A *worker* can *constitute* or *fall within* a group (being treated as an element of a group not as an individual), *join* a group, *signify* and *believe* (generally speaking express his/her opinion) and *remain* or *become* somebody else (for instance a *jobholder*). A review of these particular verbs shows that the term *employee* refers to an individual enjoying a number of rights and at the same time submitted to certain rules imposed on him/her. On the contrary, the term *worker* means a person belonging to a group and treated as its component – verbs indicating such a meaning prevail in terms of frequency and form the strongest collocations. The fact that *workers* are also individuals is (seemingly) secondary as the use of verbs shows. This is additionally stressed by the observation that a *worker* is not said to die, only the *employee* is.

The very last group of collocates with *employee* and *worker's* "possessions" shows that both have a *contract* and *employment*. Yet, only an *employee* has an *entitlement* and *right* to something. Another difference is the type of payment; for example, an *employee* receives *remuneration* and a *worker*, *wages*.

To sum up, the logDice score and absolute frequency results obtained with the word sketch feature brought interesting results in terms of multi-word terms and specialised collocations of *employee* and *worker*. The noun collocates do not contribute significantly to the meaning of

22 Worth comparing is Michta and Mroczyńska's dictionary of legal English collocations based on the UK Supreme Court judgements (Michta & Mroczyńska, 2022), where the term *employee* is a separate entry (one of 100 included therein). Despite the fact that the dictionary is based on US law, the list of collocates of *employee* coincides to a large extent with the one drawn up here, which indicates the typicality and strength of the word combinations.

the nodes. They appear to be a source of multi-word terms nesting the studied terms. However, verb collocates illustrate the meanings of *employee* and *worker* relatively well as they demonstrate what can be done to an employee and worker and what an employee and worker can do. Based on this analysis it seems reasonable to conclude that the shared meaning of *employee* and *worker* is reflected in the use of the same verb collocates, and where the collocates differ, the meaning of the terms may differ in that scope as well.

Table 4 presents the collocates of *jobholder*, *agent*, *staff* and *workforce*, which form much less representative groups than those of *employee* and *worker* owing to the low frequency of those words (see Table 2). The frequency of individual collocations formed by those lexical units is also relatively low.

Table 4. Collocates of jobholder, agent, staff and workforce in the corpus

NODE	JOBHOLDER	AGENT	STAFF	WORKFORCE
premodifier + NODE	relevant	estate	House of Commons, House of Lords, Speaker's personal, part-time, requisite, Parliamentary, academic, prison, mobile, full-time, comparable	-
NODE + noun	-	-	cost, overheads, negotiations, references	agreement
verb + NODE[23]	induce, become	authorise	appoint, recruit, exist, exclude, employ	entitle
NODE + verb[24]	become, cease, opt out of, remain, pay, authorise	-	restrict, provide, include	
NODE'S + noun	request membership, employer, right	-	Speaker, officer	-

The contextual relations in which *jobholder* is found are limited. There are no noun collocates apart from those where *jobholder* is followed with the Saxon Genitive. This specific group of collocates illustrates a relation to the terms *employee* and *worker* (the collocate *right* is shared with *employee* and the collocate *employer* with *worker*). Based on the verb collocates one can

23 The verb *be* has been excluded from the list as being too common to form a typical collocation without an additional object that would make it specific.

24 The verbs *be*, *have*, *give* and *do* have been excluded from the list as being too common to form a typical collocation without an additional object that would make them specific.

deduce that certain aspects of a *jobholder's* existence depend on his/her will (*authorise, opt out of*) and some are imposed on a *jobholder* (*cease, remain*).

As for the term *agent*, the corpus provides even fewer examples. Indeed, there is only one noun collocate *estate* and one verb collocate *authorise*. The former together with *agent* form a legal term defined in the Equality Act 2010.

Similarly, few collocates are listed under *workforce*, but still some similarities with *employee* and *worker* are evident: the fact that there is an *agreement* to which *workforce* is a party, and that *workforce* is *entitled* to something. In fact, a *workforce agreement* is a legal term in the corpus and means an agreement between an employer and his employees or their representatives. The definition implies that *workforce* is a group of employees and shows the semantic relation between this term and *employee*.

Staff, on the other hand, takes on quite a specific meaning based on the context provided in the corpus. A look at the premodifiers of *staff* shows that it primarily relates to the workers of the UK Parliament, including its two houses and the Speaker. Thus, *staff* directly relates to a particular organisation as defined above. *Part-time* and the collocates listed further in that part of the table show low absolute frequency (under 3). Yet the fact that such collocates as *part-time*, *mobile*, *full-time* and *comparable* are used with *staff* indicates that the term shares some characteristics with *employee* and *worker*, as the same collocates are also used with *employee* and *worker*. The legislator's verb selection suggests that *staff* is subordinate to another entity that can *appoint*, *recruit* or *employ* it (in other words bring into existence) and also can *exclude staff* from something.

6. Findings and conclusions

While there have been a number of corpus-based studies on legal terminologies, synonymy in legal language and collocations in legal discourse, this study focuses on the field of law that has not been in the centre of attention in terminological research so far. Employment law varies depending on the legal system and the country in which it applies. The UK legal language of employment law has its special features, including its typical terminology. The objective of this study was to empirically verify whether the terms *employee*, *worker*, *jobholder, agent*, *staff* and *workforce* are synonyms based on their dictionary and legal definitions, and whether their distant and immediate context of use can help differentiate between them.

The findings show that the context and co-text provide significant insights into the meaning of the terms under analysis. Although the sample corpus is of moderate size, the word combinations found in it allow us to draw certain conclusions as to the meaning of individual terms and the relations between them.

The study of the distribution of these terms suggests that they are applied in different areas of employment law, and thus the terms are not interchangeable. They are unevenly distributed over the 12 statutes. First of all, the frequency of the terms differs: *employee* and *worker* appear most often, while the remaining four are less frequent: from 175 (*jobholder*) to 47 (*agent*). The

distribution of the terms varies as well: *jobholder* is used only in one act, which shows that its applicability is limited to that area of law; *workforce* is found in six statutes but in three it is much more frequent than in others; *employee* is present in all the legislative documents, but its frequency in each varies and the variability is not proportional to the document size. In some cases, *worker* seems to be preferred over *employee*; *agent* and *staff* are scattered around the whole corpus and do not appear only in one act, where only *employee* and *workforce* are used.

The dictionary definitions of the terms from legal dictionaries, a business dictionary and a general English dictionary as well as the legal definitions found in the 12 statutes indicate that the terms are partial synonyms, spread along the scale of synonymity. The closest synonyms are *employee* and *worker*, but the legislator did not use them interchangeably. The other four, *jobholder*, *agent*, *staff* and *workforce*, are more distant in meaning from *employee* than *worker*, but still they semantically overlap, as demonstrated above. *Agent* appears to be the most distant in meaning, based on its dictionary and legal definitions.

The list of words (nouns, verbs and adjectives) that most frequently collocate with the terms, as identified with corpus analysis tools, has proven useful for understanding the differences between the terms under analysis. In fact, the collocational profiles of the terms seem to reflect at least part of their dictionary meanings. However, not all collocates contribute to that end equally. Most information can be deducted from combinations with verbs, which explicitly indicate what an employee, worker, jobholder, agent, staff and workforce can do and what can be done to them. As a result, it is possible to see elements of the definitions of the terms in their verb collocates. In some cases, the word combinations with the terms even extend the definitions with additional semantic information, thereby contributing to a deeper understanding of the concepts the terms represent. For example, the study of verb combinations with *employee* and *worker* showed that an *employee* is treated more as an individual, whereas a *worker* seems to be perceived as part of a larger whole.

Additionally, the study of word combinations has shown that the terms do not appear to be interchangeable based on their co-text, as the collocates they appear with differ. The extent to which their collocates overlap coincides with their shared meanings, while collocates that seem to be typical to each term suggest differences in meaning. Additionally, word combinations with nouns often form multi-word terms with the terms under analysis, which can serve as examples of nested terms. They usually stand for a specific group of employees, workers, jobholders, etc.

The study has shown that the legislative use of these terms is restricted by context, both immediate and wide, of a given paragraph, section or enactment. This study opens new avenues for exploration. It would be worth investigating how these terms behave in the secondary legal genre and whether they tend to lose at least part of the meaning assigned to them in the statutes when used in the lower-order genre. The focus could also shift towards general language, following the line of research already initiated by other scholars (L'Homme & Azoulay, 2020; Michta, 2022). This could involve comparing the specialised collocations of the six terms with those found in a general language corpus in order to learn more about the typicality of these phrases and their use and meaning in discourse outside the legal domain.

References

Adams, K.A. (2004). *A manual of style for contract drafting*. ABA Publishing, Section of Business Law.

Austin, J. (1962). *How to do things with words*. Oxford University Press.

Benson, M., Benson, E. & Ilson, R. (2009). *The BBI combinatory dictionary of English: your guide to collocations and grammar* (3rd edition). John Benjamins.

Bhatia, V. K. (2006). Legal genres. In K. Brown (Ed.), *Encyclopaedia of language and linguistics* vol. 7 (pp. 1–7). Elsevier.

Bhatia, V.K. (1993). *Analysing genre: language use in professional settings*. Longman.

Biel, Ł. (2012). Areas of similarity and difference in legal phraseology: collocations of key terms in UK and Polish company law. In A. Pamies, J. M. P. Bretaña & L. L. Nadal (Eds.), *Phraseology and discourse: cross-linguistic and corpus-based approaches* (pp. 225–233). Schneider Verlag.

Biel, Ł, & Koźbiał, D. (2020). How do translators handle (near-)synonymous legal terms? A mixed-genre parallel corpus study into the variation of EU English-Polish competition law terminology. *Estudios de Traducción, 10*, 69–90. DOI:10.5209/estr.68054.

Brezina, V. (2018). *Statistics in corpus linguistics: A practical guide*. Cambridge University Press. https://doi.org/10.1017/9781316410899.

Burkhanov, I. (1998). *Lexicography. A dictionary of basic terminology*. Wyższa Szkoła Pedagogiczna w Rzeszowie.

Buşilă, A. (2016). The issue of quasi-synonymy in the translation of notary acts from English in Romanian. *ANADISS*, *21*, 182–193.

Cao, D. (2007). *Translating law*. Multilingual Matters LTD.

Carvalho, L. (2008). Translating contracts and agreements: a corpus linguistics perspective. *Culturas Jurídicas*, *3*(1), 1–15.

Chroma, M. (2011). Synonymy and polysemy in legal terminology and their applications to bilingual and bijural translation. *Research in Language, 9*(1), 31–50.

Clark, E. V. (1992). Conventionality and contrast: Pragmatic principles with lexical consequences. In A. Lehrer, E.F. Kittay, & R. Lehrer (Eds.), *Frames, fields, and contrasts: new essays in semantic and lexical organization* (pp. 171–188). Routledge.

Cowie, A. P. (1994). Phraseology. In R.E. Asher (Ed.), *The encyclopaedia of language and linguistics* (pp. 3168–3171). Pergamon Press.

Cowie, A. P. (1998). *Phraseology. Theory, analysis, and applications*. Clarendon Press.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge University Press.

Crystal, D. (2006). *Stories of English*. Penguin.

Danet, B. (1980). Language in the courtroom. In H. Giles, W. P. Smith & P.M. Robinson (Eds.), *Language: Social and psychological perspectives* (pp. 367–376). Pergamon.

De Groot, G.-R. (1996). *Guidelines for choosing neologisms.* In: *Translation and meaning: Part 4* (pp. 377–380). Rijkshogeschool Maastricht.

Edmonds, P. & Hirst, G. (2002). Near-synonymy and lexical choice. *Computational Linguistics*, *28*(2), 105–144.

Firth, J. R. 1957/1968. A synopsis of linguistic theory, 1930–1955. In F.R. Palmer (Ed.), *Selected Papers of J.R. Firth 1952–1959* (pp. 168–205). Indiana University Press.

Goźdź-Roszkowski, S. (2012). Legal language. In C.A. Chapelle (Ed.), *Encyclopaedia of applied linguistics*. John Wiley & Sons. doi/10.1002/9781405198431.wbeal0678

Goźdź-Roszkowski, S. (2013). Exploring near-synonymous terms in legal language. A corpus-based, phraseological perspective. *Linguistica Antverpiensia, New Series – Themes in Translation Studies*, *12*, 94–109.

Halliday, M. (1961). Categories of the theory of grammar. *Word, 17*(2), 241–292.

Halliday, M.A.K. (1999). The notion of "context" in language education. In M. Ghadessy (Ed.), *Text and context in functional linguistics* (pp. 1–24). John Benjamins.

Hartmann, R. & James, G. (2002). *Dictionary of lexicography*. Routledge.

Jopek-Bosiacka, A. (2011). Defining law terms: a cross cultural perspective. In S. Goźdź-Roszkowski & I. Witczak-Plisiecka (Eds.), *Research in language. special issue on legal terminology: approaches and applications* (pp. 9–29). Łódź University Press.

Kaczmarek, K. (2013). Precision and vagueness in the language of the law in Hungarian and Polish legal texts. *Comparative Legilinguistics*, *13*, 51–68.

Klabal, O. (2019). Corpora in legal translation: overcoming terminological and phraseological asymmetries[25] between Czech and English. *CLINA*, *5*(2), 165–186.

Klabal, O. (2022). Synonyms as a challenge in legal translation training. *Białostockie Studia Prawnicze*, *27*(4), 69–82.

Kornacka, M. (2005). Termin. In J. Lukszyn (Ed.), *Języki specjalistyczne. Słownik terminologii przedmiotowej* (p. 131). Uniwersytet Warszawski.

L'Homme, M.-C. & Azoulay, D. (2020). Collecting collocations from general and specialised corpora: A comparative analysis. In G.C. Pastor & J-P. Colson (Eds.), *Computational phraseology* (pp. 151–175). John Benjamins.

Lukszyn, J. (2001). Termin i system terminologiczny w świetle praktyki terminograficznej. In J. Lukszyn (Ed.), *Metajęzyk lingwistyki. Systemowy słownik terminologii lingwistycznej* (pp. 7–25). Wydawnictwo "Aquila".

Lyons, J. (1981). *Language and linguistics. An introduction*. Cambridge University Press.

Maley, Y. (1994). The language of the law. In J. Gibbons (Ed.), *Language and the law* (pp. 11–50). Longman.

Marciniak, M. & Mykowiecka, A. (2014a). NPMI driven recognition of nested terms. In *Proceedings of the 4th International Workshop on Computational Terminology (Computerm)* (pp. 33–41). Association for Computational Linguistics and Dublin City University. DOI:10.3115/v1/W14-4805.

25 A correction to 'assymetries', the spelling used by the author of the original article.

Marciniak, M. & Mykowiecka, A. (2014b). Terminology extraction from medical texts in Polish. *Journal of Biomedical Semantics, 5*(24). DOI:10.1186/2041-1480-5-24.

Martin, E. A. (2003). *A dictionary of law* (5th edition). Oxford University Press.

Matilla, H. (2006). *Comparative legal linguistics*. Ashgate.

Matulewska, A. (2016). Semantic relations between legal terms. A case study of the intralingual relations of synonymy. *Studies in Logic, Grammar and Rhetoric, 45*(1), 161–174.

Michta, T. & Mroczyńska, K. (2022). *Towards a dictionary of legal English collocations*. Wydawnictwo Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach.

Michta, T. (2022). You shall know a term by the company it keeps: Collocations of the term *evidence* in general and legal corpora. *Beyond Philology: An International Journal of Linguistics, Literary Studies and English Language Teaching*, *19*(1), 65–96.

Murphy, M. L. (2003). *Semantic relations and the lexicon*. Cambridge University Press.

Rycak, M. B. (2008). *Wymiar i rozkład czasu pracy – monografia naukowa*. Wolters Kluwer Polska.

Rzepkowska, A. (2021). Polish-English LSP dictionaries in translation work: Labour-law terminology from the Polish Labour Code in terminographic and translation practice. *Language Culture Politics International Journal*, *1*/2021, 15–46. DOI:10.54515/lcp.2021.1.15-46

Siepmann, D. (2005). Collocation, colligation and encoding dictionaries. Part I. Lexicological aspects*. International Journal of Lexicography*, *18*(4), 409–443.

Sinclair, J. (2005). Corpus and text – basic principles. In M. Wynne (Ed.), *Developing linguistic corpora: A guide to good practice* (pp. 1–16). Oxbow Books.

Trosborg, A. (1995). Statutes and contracts: An analysis of legal speech acts in the English language of the law. *Journal of Pragmatics*, *23*(1), 31–53.

Ullmann, S. (1972). *Semantics: an introduction to the science of meaning*. Basil Blackwell.

Vázquez y del Árbol, E. I. (2006). La Traducción al español de expresiones binomiales y trinomiales (doublet & triplet expressions) en inglés jurídico: el caso de los testamentos (wills). *Babel Afial*, *15*, 19–27.

Vázquez y del Árbol, E. I. (2014). Binomios, trinomios y tetranomios cuasi sinónimos en los poderes notariales digitales británicos y norteamericanos: análisis y propuesta de traducción. *Revista de llengua i dret*, *61*, 26–46.

Webb, J. E. (2010). *The Penguin dictionary of law*. Penguin.

Wróblewski, B. (1948). *Język prawny i prawniczy* [Language of law and legal language]. Polska Akademia Umiejętności.

Yevchuk, A. (2021). Plesionyms as a vocabulary teaching tool: the case of Estonian EFL learners. *Sustainable Multilingualism*, *19*(1), 203–226.

Legal sources

Agency Workers Regulations 2010

Employment Rights Act 1996

Equality Act 2010

Health and Safety at Work etc. Act 1974
National Minimum Wage Act 1998
Part-Time Workers (Prevention of Less Favourable Treatment) Regulations 2000
Pensions Act 2008
The Fixed-term Employees (Prevention of Less Favourable Treatment) Regulations 2002
The Maternity and Parental Leave etc. Regulations 1999
Trade Union and Labour Relations (Consolidation) Act 1992
Transfer of Undertakings (Protection of Employment) Regulations 2006
Working time regulations 1998

Websites

https://dictionary.cambridge.org/ (Date of access 20 April 2023).
https://www.collinsdictionary.com/ (Date of access 20 April 2023).
https://www.gov.uk/browse/employing-people/contracts (Date of access 20 April 2023).
https://www.legislation.gov.uk (Date of access 20 April 2023).
https://dictionary.cambridge.org/thesaurus/ (Date of access 22 April 2023).
https://www.sketchengine.eu (Date of access 28 April 2023).
https://www.thesaurus.com (Date of access 22 April 2023).

***

Agnieszka Rzepkowska, PhD, is an assistant professor at the Institute of Linguistics and Literary Studies at the University of Siedlce, Poland. She earned a B.A. in Foreign Economic Relations from the Warsaw School of Economics, as well as a B.A. and M.A. in English Studies and a PhD in Applied Linguistics from the University of Warsaw. Her main area of interest is language for special purposes (LSP). She has published a number of papers and a monograph on terminology and terminography, focusing on the structure and applicability of interdisciplinary professional dictionaries. Her recent research focuses on corpus studies in LSP, particularly legal language, specialist collocations, and LSP teaching.

WORK IN PROGRESS

YURII CHYBRAS[1]
Masaryk University, Czech Republic
https://orcid.org/0000-0003-4480-7466

DOI: 10.15290/CR.2024.45.2.07

Phonetics and phonology of sound perception in a changing system

Abstract. Since the establishment of phonology as a separate branch of linguistics, scholars such as N. Trubetzkoy, C. B. Chang, E. de Leeuw, D. LaCharité, and others have demonstrated that phonological principles serve as the fundamental framework for sound perception. In particular, the key concepts of phonological sieve, approximation, language attrition and language drift show steady patterns of phonology driven sound perception. However, not all instances of sound perception adhere strictly to such phonological principles. This article examines a case of sound perception in Ukrainian revealing that, under the circumstances of phonological instability, the basic principle of sound perception may tend to shift from phonologically to phonetically driven sound perception.

Keywords: phonetics, phonology, sound perception, Ukrainian, language attrition, language drift

1. Introduction

The field of linguistics is a tapestry of interconnected facets, each contributing to our understanding of how human beings communicate through language. Among these facets, phonetics and phonology play pivotal roles, acting as gateways to unraveling the mechanisms of sound perception. In linguistics, phonetics and phonology serve as cornerstones for exploring the auditory dimension of language. Phonetics delves into the physical properties of speech sounds, examining the articulatory and acoustic features that give rise to distinct phonemes. On the other hand, phonology focuses on the more abstract, cognitive aspects of sound patterns, seeking to understand how humans categorize and process these sounds within the boundaries of their linguistic systems. This duality between phonetics and phonology contributes to the complexity of studying sound perception.

This paper explores the intersection of phonetics and phonology, tracing the history of dominant approaches to the study of sound perception and examining a case that provides valuable data to expand the current understanding of the mechanisms behind sound perception.

1 Address for correspondence: Department of English and American Studies, Faculty of Arts, Masaryk University, Arna Novaka 1, Brno 602 00, Czech Republic. E-mail: cibras@mail.muni.cz

The following sections will be dedicated to the history of sound perception as a subdiscipline within phonetics, the formation of the current approach, and possible ways to further develop the dominant framework, as illustrated through an experimental study.

2. History

Since its establishment as a significant part of linguistics, phonology has made substantial progress thanks to the efforts of many prominent linguists. In *Grundzüge der Phonologie*, N. Trubetzkoy defined and formalized what phonology is, and how it differs from phonetics and other related fields. One of the key concepts introduced at the beginning of *Grundzüge* is the notion of the "phonological sieve", described by Trubetzkoy as:

> The phonological system of a language is like a sieve through which everything that is said passes … Each person acquires the system of his mother tongue. But when he hears another language spoken, he intuitively uses the familiar 'phonological sieve' of his mother tongue to analyze what has been said. However, since this sieve is not suited for the foreign language, numerous mistakes and misinterpretations are the result. The sounds of the foreign language receive an incorrect phonological interpretation since they are strained through the 'phonological sieve' of one's own mother tongue. (Trubetzkoy, 1939, p. 51).

In this paragraph, Trubetzkoy essentially describes the foundational principle of sound perception. According to him, perception follows the rules of phonology, meaning that phonetic similarity plays a lesser role in how sounds are perceived.

Later, in the United States during the 1950s and 1960s, the topic of sound perception, along with sound production, gained renewed interest among linguists. With several new waves of migration from Europe and other countries after World War II, and thousands of first-generation immigrants learning English, linguists became interested in *system-in-change*. This was the period when the field of second language acquisition, along with other related phonetic and phonological aspects of non-native speech, gained prominence. Concepts such as *phonological interference*, *phonological variability*, and the *phonetic and phonological manifestations of L1 phonology*, among others, entered the linguistic discourse of the time.

One of the first works to address a related topic appears to be William Nemser's *Approximative Systems of Foreign Language Learners* (1971). Although Nemser did not specifically focus on sound perception in this work, one of the issues he studied was sound perception within a system experiencing rapid change. The outcome was a concept that extended beyond the traditional understanding of sound perception, namely, the idea of an approximative system — *the deviant linguistic system actually employed by the learner attempting to utilize the target language* (Nemser, 1971, p. 2). According to Nemser, the reason learners develop an approximative system is essentially the same as Trubetzkoy's notion of a phonological sieve – the speaker's inability to properly

hear and produce a new set of phonemes. Nemser's concept also expanded upon Trubetzkoy's originally phonological idea by incorporating semantics, syntax, and other linguistic areas.

This approximative system, according to Nemser, is an ever-evolving system that undergoes abnormally rapid changes as the learner progresses. The system is also coherent and forms a patterned product distinct from both the learner's L_1 and L_2 systems (in the original work marked as LS, LT, and La, standing for the *source language*, *target language*, and *approximative system*, respectively).

This period also saw the emergence of similar new concepts. For instance, in his 1980 work Phonetic Approximation in Second Language Acquisition, James E. Flege first introduced the concept of phonetic approximation. In this study, Flege researched how the L_1 phonology of Saudi Arabic speakers manifests in their L2 English output. He concluded that the more experienced Saudi English speakers produced word-final stops closer to those found in native American English while still maintaining phonological features typical of Saudi Arabic. This resulted in a sort of intermediate (or approximative) system.

As the new millennium approached, linguists continued to explore principles of sound perception. Starting in 1995, linguists from Quebec's Laval University, such as Darlene LaCharité, Carole Paradis, and their colleagues, conducted research on loanword phonetic approximation – that is, how English loanwords were adapted into the system of Quebecois French. Their research topics include LaCharité & Paradis (2002), LaCharité & Prévost (1999), and Paradis & LaCharité (2008). In Poznań, Poland, Ewa Waniek-Klimczak and her colleagues have worked on areas related to second-language pronunciation (Waniek-Klimczak, 2009, 2014, 2016, 2019; Waniek-Klimczak et al., 2015; Waniek-Klimczak & Shockey, 2013), focusing primarily on the speech of Polish immigrants and the Polish accent in English as a second language, as well as its perception. The Linguistics Department at Boston University has produced numerous works in the field of phonetics and phonology, contributing valuable research to the discussion on the role of phonetics and phonology in non-native, second/third/heritage speech, language attrition, and drift (Chang, 2008, 2013, 2019a, 2019b; Chang & Ahn, 2023; Chang & Dionne, 2022; Chang & Kwon, 2020; Hutchinson, 2022).

The conclusion of the studies mentioned before is that the main principle of sound perception follows the principles of phonology. The phonetic aspect, however, seems to manifest mainly in sound production. Thus, for instance, Flege described *phonetic approximation* as a phenomenon when a speaker is unable to accurately produce the correct target L2 sound and produces a (range of) rough approximation(s) instead:

> In much previous research, especially that done within a phonemic theory framework the L2 sounds produced by a language learner have often been viewed as discrete entities which are produced either correctly or incorrectly instead of as a continuum of approximations to phonetically accurate L2 sounds. (Flege, 1980, p. 119) [emphasis added]

or as in:

> Importantly, it seemed to be the case that the least correct mispronunciations tended to disappear first from the learner's speech, while the closer (but still phonetically inaccurate) approximations to L2 phonemes remained longer. (Flege, 1980, p. 119) [emphasis added].

To put it differently, phonology is concerned with sound perception, while phonetic approximation accounts for a range of unstable acoustic realizations. Nevertheless, some authors suggest that such influence may also extend to the phonetic level. For instance, Al-Kinany et al. (2022), Hasan et al. (2011), Kodirova (2021), Rohali (2018), and Soares (2012) mention *phonetic interference* as the influence of L1 on the L2 production by non-native speakers. Despite varying terminology, their ideas reflect a common theme: *phonology*-driven *phonetic* realization. This is closely related to articulatory variability, a phenomenon where speakers struggle to accurately produce the target L2 sound, instead generating a range of rough approximations, as described by Yun and Sung (2022). Thus far, instability has been discussed as a cause of phonetic variation in speech. This raises the question of whether such instability can also lead to variation in sound perception.

3. Phonetics or phonology?

Proponents of the phonetic approach argue that sound perception occurs at the phonetic level, meaning sounds are perceived based on their phonetic proximity. This is often the case with bilinguals, who naturally attempt to approximate the closest possible pronunciation. This perspective was described in Paradis and LaCharité's (2008) paper on apparent phonetic approximation in Quebecois French during the 19th and 20th centuries. However, they concluded that sound perception in loanword adaptation is phonologically rather than phonetically driven. In contrast, proponents of the phonological approach argue that sounds are adapted not only based on their phonetic proximity but also in accordance with the rules of the respective phonological systems, (false) analogies, and other factors. Although a final consensus has not yet been reached, it appears that more linguists – such as Larry Hyman, Mathias Jenny, Michael Kenstowicz, Darlene LaCharité, Lynn Nichols, Donca Steriade, Bert Vaux, and Jie Zhang – lean towards the phonological approach.

The hypothesis of this article is that sound perception typically follows phonological principles when the phonological system is in a relatively stable state. However, when a phonological system is undergoing change or when a significant number of speakers experience prolonged language *attrition* (as discussed by de Leeuw and Chang, 2023), sound perception may, at least partially, follow phonetic principles. In de Leeuw and Chang's conceptualization, language attrition—contrary to the more commonly known understanding—refers to nearly permanent changes in an individual's language system, usually affecting grammar and phonology, resulting from prolonged language contact.

To gain a better practical understanding, an example should be introduced into the discussion. A case in the English-Ukrainian language pair and phonemes /g/, /ɦ/, /h/ /x/ might illustrate

the hypothesis. English possesses two out of four phonemes: /g/ and /h/ as in *growl* and *house*. The phonological inventory of Ukrainian possesses three of them: /g/, /ɦ/ and /x/, as in <ґанок> [ganɔk], <ґарний> [ɦarneĭ] and <ходити> [xɔˈdɪte]. Having such phonemes in its inventory, as well as applying the principles of phonology, one would expect Ukrainian speakers to perceive and adapt English /g/ and /h/ as /g/ and /ɦ/ respectively, as /g/ is present in both phonological systems and /h/-/ɦ/ share the same place of articulation as both phonemes are glottal fricatives. Such principles of phonology, for example, work as expected for phonologically similar Czech (Mołęda, 2008; Duběda, 2020). Czech speakers, therefore, tend to perceive /ɦ/ as the closest equivalent of /h/. The situation with Ukrainian speakers is, however, somewhat different. The phoneme /g/ in Ukrainian is rather marginal. It used to be a core part of the inventory of its ancestor language, Proto-Slavic, while at later stages of its development etymological /g/ first changed into /ɣ/, subsequently changing into /ɦ/ during the 10-13th centuries. This process also involved domesticating loanwords, e.g., the word *etymology* itself is [etemɔˈɬɔɦʲijɛ] in contemporary Ukrainian. The list of frequently used words containing the phoneme /g/ would not be longer than approximately 20 vocabulary units that, in spoken Ukrainian, *may* also be pronounced with [ɦ] instead.

Furthermore, there are two other sociolinguistic factors that should be mentioned. The first factor is that most Ukrainians are either Ukrainian-Russian bilinguals or advanced L2 speakers of Russian. They are also aware of the so-called hekannia, i.e., pronouncing [ɦ] instead of [g] when speaking Russian – a phonological feature characteristic of Russian spoken in Ukraine. Additionally, the two languages share multiple cognates where the Russian word would contain [g] while the Ukrainian word would contain [ɦ] as, for instance, is the case with Russian гора [gɐˈra] and Ukrainian гора [ɦɔˈra]. Such words are easily recognisable and oftentimes have (nearly) identical spelling. This sociolinguistic feature has created a strong link between the two sounds and, as a result, it influences how Ukrainian speakers tend to render /g/ and [h]-like sounds of foreign languages. Eventually, it creates a phonological collision when perceiving and adapting English /g/ and /h/. A speaker indicates that growl and house have different initial consonants in the original language. To avoid a collision the latter is often rendered as [x] instead, thus giving us [ɦrɔuɬ] and [xaus] respectively.

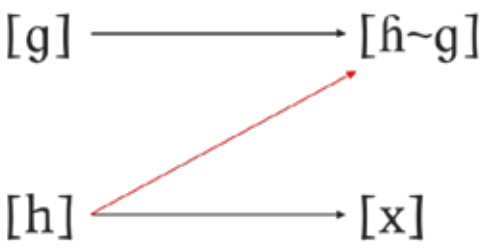

Figure 1. Phonetically driven [x]-based perception (Chybras, 2024). The red line indicates the phonologically expected perception.

The second sociolinguistic factor is orthography. The latest Ukrainian orthography from 2019 (The Cabinet of Ministers of Ukraine, 2019) states that foreign /g/ and /h/ sounds are to be transliterated as <r>, which stands for /ɦ/, while there are certain exceptions where a foreign, often English, /h/ may be transliterated as <x>, which stands for /x/. In most, if not all, cases this tradition of transliteration is Russian influenced as Russian used to be a mediator language, i.e., loanwords would enter Ukrainian through Russian. For instance, the classical orthography of Ukrainian from 1928 mentioned that foreign /h/ should be adapted as <r> while foreign /g/ should be adapted as <ґ>, [g] phonetically, except for the older loanwords, that is, those that had undergone the process of a /g/ to /ɦ/ change (Skrypnyk, 1929). Additionally, the same 2019 orthography update mentions that in personal names <g> and /g/ *may* be transliterated as <ґ>, which stands for /g/. As the majority of contemporary loanwords are of English origin, Ukrainian speakers tend to apply this exception to newly borrowed words usually stating *phonetic similarity* as the reason for choosing <x>. This state of affairs has also led to decades-long discussions about correct transliteration methods (Molotkina, 2017).

This phonological anomaly was the primary research interest in Chybras (2024). The study aimed to reveal why some Ukrainian speakers tend to have [x]-based perception of foreign [h] instead of the phonologically expected [ɦ]-based perception, as previously described in the hypothesis. The results obtained from the study showed there is a correlation between exposure to Ukrainian-accented Russian and [x]-based perception of [h]. The most prone to [x]-based perception group proved to be the bilingual group the phonological system of which has experienced prolonged *attrition* (as in de Leeuw & Chang, 2023). Their phonological system of Ukrainian, therefore, seems to exist in a state of instability and change. This phonological instability, thus, creates favourable conditions for phonetic and phonological uncertainty in sound perception. That is when two sounds, [g] and [ɦ], are ultimately perceived as one phoneme with peculiar realisation patterns, meaning that [g] can be realised as [ɦ], while [ɦ] cannot be realised as [g], therefore [h] cannot be perceived as [ɦ], the phonetically and phonologically closest sound.

The study was conducted as a single-blind experiment that studied the participants' phonetic perception. A total of 34 participants were asked to listen to the recordings of words containing [g] and [h] in various positions and combinations. The null hypothesis of the study was that the reason why some Ukrainian speakers perceive [h] as [x] is purely phonetic, and Czech [ɦ]-based perception is caused by orthography. This, however, was disproved, with the ultimate reason identified as language attrition, which affects the phonological system of bilingual speakers of Ukrainian living in a mixed language environment. Furthermore, respondents from the bilingual group exhibited mixed perceptions of [h], that is, in one phonetic environment, they would perceive [h] as [x] while in a different phonetic environment they would perceive [h] as [ɦ]. Interestingly, they tended to perceive [h] as [ɦ] when [x] was also present in the neighbouring position in a word, thus contrasting the two sounds. The results showed that the bilingual group, on average, reached $\bar{x}$ 80.1 % [x]-based perception, while the results for the Ukrainian-dominant and the Russian-dominant groups were $\bar{x}$ 56.2 % and $\bar{x}$ 69.9 % respectively.

Furthermore, the results suggest that there might be positional tendencies in sound perception among Ukrainian speakers. Although the position did not appear to be the original cause of [x]-based perception, it seems plausible to argue that such tendencies arise from phonological instability. For instance, it was revealed that the respondents tended to perceive [h] as [x] in positions such as [VhV] and [hV̥Ç] where the perception of [h] was either contrasted to vowels or influenced by regressive assimilation by a following voiceless consonant. Contrary to that, [h] tended to be perceived as [ɦ] in positions like [Çh], [hVÇ] and [xVhV]. These positions either contain a voiced consonant close to [h] or [h] is contrasted to [x].

4. Implications

These findings establish a foundational premise for the ensuing argument: the impact of phonology on sound perception diminishes when a language system undergoes changes in its phonological structure, or when specific phonemes within that system occupy peripheral and precarious positions. In such scenarios, phonetically driven perception becomes more reliable from the perspective of language speakers. However, the uniformity of this perceptual shift across diverse languages requires further scholarly inquiry.

Currently, several conceptual results can be discerned. While the concepts of *articulatory variability* and *phonetic approximation* primarily characterize phenomena in the domain of speech production, *perceptual variability* – manifested in speech perception – seems to arise from phonological instability. Additionally, the structural foundations of phonology-driven perception appear to transition to a phonetics-driven paradigm, often characterized by an ad hoc and functional perceptual approach influenced by the surrounding phonetic context. This suggests that in an unstable phonological environment, sound perception may revert to a lower (phonetic) structure, which is more easily comprehensible, rather than a higher (phonological) structure that demands stronger connections and oppositions between phonemes. If these observations withstand scrutiny, as supported by similar phenomena in other language combinations, they should enhance our understanding of sound perception.

References

Al-Kinany, T., Al-Abri, A., & Ambusaidi, H. (2022). Arab EFL learner perceptions of English phonemes: A cross-language phonetic interference. *English Language Teaching*, *15*(2), 67. https://doi.org/10.5539/ELT.V15N2P67

Brown-Bousfield, M. & Chang, C. (2023). *Regressive cross-linguistic influence in multilingual speech rhythm: The role of language similarity. L3 Development after the initial state*. John Benjamins. DOI: 10.1075/sibil.65.03bro

Chang, C. B. (2008). Phonetics vs. phonology in loanword adaptation: Revisiting the role of the bilingual. *Annual Meeting of the Berkeley Linguistics Society*, *34*(1), 61. https://doi.org/10.3765/bls.v34i1.3557

Chang, C.B. (2010). First language phonetic drift during second language acquisition. [Unpublished doctoral dissertation]. UC Berkeley.

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, *41*(6), 520–533. https://doi.org/10.1016/J.WOCN.2013.09.006

Chang, C. B. (2019a). The phonetics of second language learning and bilingualism. *The Routledge handbook of phonetics* (pp. 427–447). Routledge. https://doi.org/10.4324/9780429056253-16

Chang, C. B. (2019b). Phonetic drift. *The Oxford handbook of language attrition* (pp. 190–203). Oxford University Press. https://doi.org/10.1093/OXFORDHB/9780198793595.013.16

Chang, C. B. (2021). Phonetics and phonology of heritage languages. *The Cambridge handbook of heritage languages and linguistics, Chapter 23.* Cambridge University Press.

Chang, C. B., & Ahn, S. (2023). Examining the role of phoneme frequency in first language perceptual attrition. *Languages*, *8*(1), 53. https://doi.org/10.3390/LANGUAGES8010053

Chang, C. B., & Dionne, D. (2022). Unity and diversity in Asian American language variation: Data from Chinese, Filipino, Korean, and Vietnamese Americans. *Fourth Vienna Talk on Music Acoustics*, *49*, 060002. https://doi.org/10.1121/2.0001669

Chang, C. B., & Kwon, S. (2020). The contributions of crosslinguistic influence and individual differences to nonnative speech perception. *Languages*, *5*(4), 1–27. https://doi.org/10.3390/LANGUAGES5040049

Chybras, Y. (2024). Phonology and attrition, sociolinguistics of (Ukrainian) sound perception. *Półrocznik Językoznawczy Tertium*, *8*(2), 147–164. DOI: 10.7592/Tertium.2023.8.2.259

De Leeuw, E. & Chang, C. B. (2023). Phonetic and phonological L1 attrition and drift in bilingual speech. *The Cambridge handbook of bilingual phonetics and phonology*. Cambridge University Press.

Duběda, T. (2020). The phonology of anglicisms in French, German and Czech: A contrastive approach. *Journal of Language Contact*, *13*(2), 327–350. https://doi.org/10.1163/19552629-01302003

Flege, J. E. (1980). Phonetic approximation in second language acquisition. *Language Learning*, *30*(1), 117–134. https://doi.org/10.1111/j.1467-1770.1980.tb00154.x

Hasan, A. S., Al-Khateeb, S., & Azzouz, A. (2011). The study of lexical and phonetic interference from Arabic into English for Syrian university students. *Research Journal of Aleppo University.* https://www.researchgate.net/publication/366790113_The_Study_of_Lexical_and_Phonetic_Interference_from_Arabic_into_English_for_Syrian_University_Students

Hutchinson, A. E. (2022). Individual variability and the effect of personality on non-native speech shadowing. *JASA Express Letters*, *2*(6), 065203. https://doi.org/10.1121/10.0011753

Kodirova, K. S. (2021). Phonetic interference – As a result of the interaction of languages. *Current Research Journal of Philological Sciences*, *02*(10), 63–66. https://doi.org/10.37547/PHILOLOGICAL-CRJPS-02-10-13

LaCharité, D., & Paradis, C. (2002). Addressing and disconfirming some predictions of phonetic approximation for loanword adaptation. *Langues Et Linguistique*, *28*, 71–91.

LaCharité, D., & Prévost, P. (1999). The role of L1 and teaching in the acquisition of English sounds by francophones. *Proceedings of BUCLD* 23, 373–385.

Mołęda, J. (2008). Phonological adaptations of anglicisms in Polish and Czech. A critical view. *Bohemistyka*, 1–4, 295–308.

Molotkina, Y. (2017). Rizni sposoby napysannia novitnikh anhlitsyzmiv v ukrainskii movi. *Research Journal of Drohobych Ivan Franko Pedagogical University. Series "Philology" (Linguistics)*, *7*, 119–123.

Nemser, W. (1971). Approximative systems of foreign language learners. *IRAL*, *9*, 115–123. https://doi.org/10.1515/iral.1971.9.2.115

Paradis, C., & LaCharité, D. (2008). Apparent phonetic approximation: English loanwords in Old Quebec French. *Journal of Linguistics*, *44*(1), 87–128. https://doi.org/10.1017/S0022226707004963

Rohali, R. (2018). The phonetic interference: Indonesian into French. *4th International Symposium on Language and Arts Education (ISOLA 2018) At: Kelantan, Malaysia,* 1-8.

Skrypnyk, M. (1929). Ukrainskyi pravopys. Derzhavne Vydavnytstvo Ukrainy. http://irbis-nbuv.gov.ua/cgi-bin/ua/elib.exe?Z21ID=&I21DBN=UKRLIB&P21DBN=UKRLIB&S21STN=1&S21REF=10&S-21FMT=online_book&C21COM=S&S21CNR=20&S21P01=0&S21P02=0&S21P03=FF=&S21STR=ukr0004164

Soares, V. H. M. (2012). L1 Brazilian Portuguese phonetic interference on L2 English: An analysis based on corpora. [Unpublished Master's thesis]. Universidade Federal de Minas Gerais. https://doi.org/10.13140/RG.2.2.28531.55845

The Cabinet of Ministers of Ukraine. (2019). Ukrainian orthography. https://mon.gov.ua/storage/app/media/zagalna%20serednya/%202019.pdf

Trubetzkoy, N. S. (1939). *Grundzüge der Phonologie*. Akciová moravská knihtiskárna Polygrafie v Brně.

Waniek-Klimczak, E. (2009). Sociolinguistic conditioning of phonetic category realisation in non-native speech. *Research in Language*, *7*, 149–166. https://doi.org/10.2478/V10015-009-0010-9

Waniek-Klimczak, E. (2014). Selected observations on the effect of rhythm on proficiency, accuracy and fluency in non-native English speech. *Second Language Learning and Teaching*, *19*, 167–181. https://doi.org/10.1007/978-3-319-00419-8_12

Waniek-Klimczak, E. (2016). Review of pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research; Authors: Tracey M. Derwing and Murray J. Munro; Publisher: John Benjamins, 2015; ISBN: 9789027213273; Pages: 208. *Studies in Second Language Learning and Teaching*, *6*(2), 345–348. https://doi.org/10.14746/SSLLT.2016.6.2.9

Waniek-Klimczak, E. (2019). Variable rhoticity in the speech of Polish immigrants to England. *Approaches to the study of sound structure and speech: Interdisciplinary work in honour of Katarzyna Dziubalska-Kołaczyk* (pp. 326–337). Routledge. https://doi.org/10.4324/9780429321757

Waniek-Klimczak, E., Rojczyk, A., & Porzuczek, A. (2015). Polglish in Polish eyes: What English studies majors think about their pronunciation in English. *Second Language Learning and Teaching*, *24*, 23–34. https://doi.org/10.1007/978-3-319-11092-9_2

Waniek-Klimczak, E., & Shockey, L. (2013). *Teaching and researching English accents in native and non-native speakers*. Springer.

Yun, G., & Sung, J. H. (2022). Articulatory variability of phonological rules by Korean EFL and Indian ESL speakers. *Korean Journal of English Language and Linguistics*, *22*, 1465–1493. https://doi.org/10.15738/KJELL.22..202212.1465

***

Yurii Chybras is a doctoral student at the Department of English and American studies at Masaryk University, Brno, Czech Republic. His research interests lie primarily in the fields of phonetics, phonology, sociolinguistics, and non-native English.