Table of Contents

New Developments in English for Special Purposes (ESP) Lexicology and Lexicography

Daniele Franceschi[1] Laura Pinnavaia[2]

[1] Department of Foreign Languages, Literatures and Cultures, 'Roma Tre' University, Rome, Italy

Correspondence: Daniele Franceschi, Department of Foreign Languages, Literatures and Cultures, 'Roma Tre' University, Via Ostiense 234-236, 00146 Rome, Italy. E-mail: daniele.franceschi@uniroma3.it

[1] Department of Languages, Literatures, Cultures and Mediations, University of Milan, Milan, Italy

Correspondence: Laura Pinnavaia, Department of Languages, Literatures, Cultures and Mediations, University of Milan, Piazza Sant'Alessandro 1, 20122 Milano, Italy. E-mail: laura.pinnavaia@unimi.it

Introduction

The fact that languages are living organisms that evolve and vary as a result of different factors, including societal vicissitudes, cultural influences, technological advancements, migration and intercultural communication, among others, requires a continued effort to identify, characterize and represent the new lexical forms and meanings that emerge in various contexts and domains. Lexicologists and lexicographers thus need to always take on new challenges in their attempt to adequately describe and map our ever-changing language landscapes. Not only does everyday language change but also languages for special purposes (LSP) are subject to variation. On the one hand, new terms and senses are incorporated into already existing lexicons, while, on the other hand, micro-languages with a rich variety of novel lexical items keep appearing. Some cases in point are ecotourism discourse which has been accelerated by recent greater awareness of the impact of human activities on the environment (Penz & Fill, 2022); the language of digital technology, now that societies and technologies have become increasingly interconnected (Würschinger, 2021); certain medical talk, as a consequence of the coronavirus pandemic (Salazar & Wild, 2022), to mention just a few areas of lexical innovation.

Predicting the degree of future stability of new lexical forms and meanings is no easy feat. Some of them, usually the most creative or controversial ones, typically emerge to satisfy momentary communicative needs, attitudes and intentions. Consider, for instance, the expression *eco-Nazi* recently used to refer to a very radical environmentalist, intolerant towards people with different views. Although it is represented in some of the largest corpora of English (there are 9 occurrences in the *News on the Web Corpus*, 14 in *The intelligent Web-based Corpus* and 10 in the *Corpus of Global Web-based English*) (Note 1), the expression has not found its way into dictionaries and may thus lose ground and then disappear or perhaps just continue to marginally exist without attaining lexicographic status. The opposite scenario however may also be possible, i.e., through repeated use, it may eventually appear in dictionaries. After all, the *Oxford English Dictionary* (OED) has already included morphologically similar formations, such as *neo-Nazi* and *pro-Nazi*, that emerged further to the first attestations of *Nazi* in the 1930s. On the other hand, there are new words and expressions that have gained force over the years and have thus officially been added as lexicographic entries. Still in the context of ecological discourse, the compounds *dark green* or *deep green* and *light green* are now fully lexicalized and present in the OED: a *dark/deep green* is 'a person who is deeply committed to environmental causes or who holds radical green views' as opposed to a *light green* who is instead only 'moderately or superficially committed to environmental causes, or holds moderate green views' (Note 2).

Historical dictionaries like the OED allow the study of meaning evolution retrospectively, but they cannot tell us what the future of the various lexical items listed in them will be like. Corpora instead provide useful indications about the behavior of words, whose mere frequency may suggest that they are either gaining a firm footing or declining in usage. Artificial intelligence today also represents an important resource able to support lexicologists and lexicographers in their research. With a simple query on ChatGPT (Note 3), we can quickly obtain useful neologisms in various fields that are often not yet present in corpora. AI allows us to perform generic searches, while corpora are best queried for specific words or phrases, whose behavior in context may be observed and compared with that of other ones. If we ask ChatGPT to give us examples of, say, medical neologisms, it will return very interesting data that would be difficult to retrieve otherwise, e.g. *WFH (Work From Home) syndrome*, *pandemic fatigue*, *zoom gloom* and *maskne*, which emerged during the Covid-19 pandemic (Note 4). Although not even ChatGPT can always provide examples of the very latest terms used in a certain field or discipline, it certainly represents a useful tool to collect new items and create initial glossaries. It seems indeed likely that both lexicology and lexicography will benefit greatly from generative AI in the future. This has incidentally been confirmed by

research presented at conferences in the course of 2023 and already published in mainstream high-quality journals (cf. Curry, Baker & Brookes, 2024; de Schryver, 2023; de Schryver & Joffe, 2023; Lew, 2023).

This Special Issue collects 9 research papers covering a wide range of languages for special purposes. The concept of 'specialized discourse' is broadly interpreted in that it is made to refer not only to those language contexts where jargon or technical terms are used, but also to situations where less complex but still discipline-specific vocabulary, not so common elsewhere, occurs and recurs. In other words, ESP is understood here as the specific use of English in given fields and related sub-fields, such as tourism and eco-tourism, medicine and the language of Covid-19, etc., encompassing a mixture of terminology and more understandable vocabulary, even to a non-expert ear. As a matter of fact, the degree of technicality of meaning and language choice always ranges along a continuum from precise and specialized to looser vocabulary related to various topics. In natural language the most likely scenario is to come across words, phrases and expressions that even in the same text or context exhibit different levels of complexity and comprehensibility. The former is the cause of the latter, which is strictly related to the concept of difficulty. Put differently, complexity pertains to the objective features of a certain item, while difficulty is subjective and dependent on individual factors. This is a necessary precision in that the two notions of complexity and difficulty are often confused and the two words are wrongly used as synonyms. It is instead obvious that specialized terms are usually unintelligible to the layperson, but familiar to a specialist in a certain field.

The present volume opens with Elisa Mattiello's paper which examines the role played by new combining forms (CFs) in the expansion of ESP lexicon in several areas, such as science, information technology, economics, chemistry, medicine and so forth. Because of its cross-domain orientation, it works well as an introductory and overview work on the latest neologisms of this type. The study is both qualitative and quantitative in that it discusses the addition of CFs to the OED since the second half of the last century, but it also considers their frequency and distribution in two main corpora of contemporary English, i.e. the *Corpus of Contemporary American English* (COCA) and the *News on the Web Corpus* (NOW), in order to make claims about their stability and profitability.

The following papers focus on specific ESP areas: some of them take a lexicological or lexicological-lexicographic approach, while others address strictly lexicographic issues, such as the creation of ESP glossaries, the improvement of ESP dictionary entries and the addition of lemmas to online specialized reference tools. This second group of papers include Daniele Franceschi's contribution on English-Italian dictionaries of sports and games terminology; Antonella Luporini's study of artistic vocabulary still from a bilingual perspective involving English and Italian; Gesuato, Castello and Gaballo's propose the design of a multilingual hairstyling glossary; and finally, Silvia Cacchiani's paper discusses the challenges of representing paradigmatically related Covid-19 terms in online resources.

After providing an initial overview of the existing bilingual English-Italian dictionaries of sports and games terminology, Franceschi presents an improved hypothetical e-dictionary entry to illustrate the advantages of online multimodal and multimedial lexicographic resources, while also discussing the challenges of creating and updating them. Luporini instead shows how a specialized corpus of bilingual texts specifically related to art and cultural heritage can be effectively utilized for the analysis of Italian lemmas and their English counterparts in this field; the patterns retrieved in the corpus are then compared and contrasted with the ones appearing in four general Italian-English dictionaries and the possibility of creating a large-scale specialized bilingual dictionary of art, which is non-existent to date, is discussed. Gesuato, Castello and Gaballo illustrate a corpus-driven term extraction and description procedure, focusing in particular on translation equivalents and phraseologies in English and Italian, with the aim of creating glossary entries that cater to diverse users, namely (trainee) language practitioners and professionals alike. In this first group of papers, Cacchiani contributes to the recent strand of work on the language of Covid-19 first through the examination of the glossaries available on the websites of the UK Parliament and Government, i.e. on credible and authoritative platforms, and then by comparing them with other lexicographic resources with specific reference to some selected entries.

The other four papers in the volume follow a more lexicological approach. Cristiano Furiassi examines the use of idiomatic toponymic phrasemes in Bajan, the English-based creole spoken in Barbados. The data analyzed are specifically from an amateur printed collection of expressions employing territorial place names, namely the ones of former sugar-cane plantations. It is shown how the paremiological inventory of Bajan consists not only of well-established and brand-new proverbs and proverb-like phrases, but also of lexical items that are the result of a "resemanticization" process of English-derived toponyms on the basis of the islanders' communicative practices. The latter reflect the archetypally different weltanschauung of Barbadians from that superimposed by British colonizers in the past. Silvia Sperti investigates the use of terms in the context of migration discourse typically involving English as a Lingua Franca (ELF) speakers, i.e. officers, legal advisors, mediators, asylum seekers and refugees, who contribute to the development of linguacultural representations partly deviating from those produced by native English speakers. These instantiations cannot be included in mainstream dictionaries and glossaries, which are consequently unable to satisfy ELF speakers' practical needs. Therefore, the creation of adapted

lexicographic tools, that employ a descriptive rather than a prescriptive approach, is called for. Sara Corrizzato describes some cases of lexical adaptation and innovation in promotional discourse appearing on the American Instagram pages of nine agri-food Italian companies selling their products internationally. The analysis shows that the new words and expressions identified basically aim at reinforcing the 'Made in Italy' concept. While some of them appear to have been created "on the fly" and will most probably not gain stability, other innovative uses are expected to linger in the food lexicon. Finally, Lorenzo Buonvivere's contribution is a case study on the combining form *eco(-)lodge* in the context of 'ecotourism talk', which is examined by searching both native speakers and learners' dictionaries as well as specialized and general English corpora. While examples of usage mostly describe an *eco(-)lodge* as a type of luxury and exclusive accommodation to be found in natural – i.e., non-urban – contexts, dictionaries define them only with reference to their supposed minimal environmental impact. Therefore, there appears to be a process of semantic bleaching going on in ecotourism discourse, whereby the meaning of *eco-* is exploited to advertise a form of niche tourism, which incidentally does not always align with ecological concerns.

With this volume, we have ultimately tried to put together a collection of papers showing that the lexicon of ESP is in constant flux, which makes it necessary to always monitor it so as to advance both theoretical and empirical enquiry and to be able to accommodate new forms and meanings in the dictionary.

References

de Schryver, G.-M., 2023. Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography*, 36(4): 355–387, https://doi.org/10.1093/ijl/ecad021.

de Schryver, G.-M. & D. Joffe, 2023. *The End of Lexicography, Welcome to the Machine: On How ChatGPT Can Already Take over All of the Dictionary Maker's Tasks*. Universiteit Ghent.

Curry, N. Baker, P., & G. Brookes, 2024. Generative AI for corpus approaches to discourse studies: A critical evaluation of ChatGPT. *Applied Corpus Linguistics* 4(1): 100082, https://doi.org/10.1016/j.acorp.2023.100082.

Lew, R., 2023. ChatGPT as a COBUILD lexicographer. *Humanities and Social Sciences Communication* 10, Article number 704. https://doi.org/10.1057/s41599-023-02119-6.

Penz, H. and A. Fill, 2022. Ecolinguistics: History, today, and tomorrow. *Journal of World Languages* 8(2): 232-253.

Salazar, D. and K. Wild, 2022. The Oxford English Dictionary and the language of Covid-19. In Klosa-Kückelhaus, A. & E. Kernerman (Eds), *Lexicography of Coronavirus-related neologisms* (pp. 11-26). Berlin/Boston: De Gruyter.

Würschinger, Q., 2021. Social Networks of Lexical Innovation. Investigating the Social Dynamics of Diffusion of Neologisms on Twitter. *Frontiers in Artificial Intelligence* 4: Article 648583.

Notes

Note 1. Consult https://www.english-corpora.org to assess these corpora.

Note 2. https://www.oed.com (Last access: 20 December 2023).

Note 3. https://chat.openai.com (Last access: 20 December 2023).

Note 4. People who suffer from the *WFH syndrome* typically multitask work and home chores to the point where they never get sufficient time to relax thus feeling depression, anxiety, stress, burnout and fatigue all mixed into one; the term *pandemic fatigue* was introduced by the World Health Organization (WHO) to indicate a widespread sense of distress affecting the entire population as a consequence of the prolonged state of crisis caused by the Covid-19 pandemic, whose end could not be foreseen; *zoom gloom* refers to the mental and physical exhaustion after working many hours on the Zoom platform; and *maskne* is a blend (*mask+acne*) indicating a number of skin conditions that can stem from wearing a mask.

Copyrights

The development of ESP lexicon through new combining forms

Elisa Mattiello[1]

[1] Department of Philology, Literature and Linguistics, University of Pisa, Pisa, Italy

Correspondence: Elisa Mattiello, Department of Philology, Literature and Linguistics, University of Pisa, Via Santa Maria 67, 56126 Pisa, Italy. E-mail: elisa.mattiello@unipi.it

Abstract

This paper investigates the role of new combining forms in the formation of neologisms which are currently expanding the lexicon of English for Special Purposes (ESP). In the past, only neoclassical combining forms, such as initial *bio-* or final *-logy* (in *biology*), were productively used in ESP. Nowadays, specialized combining forms also include abbreviated forms of existing words (e.g., *cyber-* from *cybernetic* in *cyber-attack*), as well as secreted (i.e. reinterpreted) forms (e.g., *-bot* from *robot* denoting 'a type of automated program or software' in *knowbot*). The paper explores a set of combining forms attested since the second half of last century in the online version of the *Oxford English Dictionary* (OED) with the aim to demonstrate how specialized sectors, such as science or information technology, are being enriched by series of combining-form combinations. The paper conducts quantitative analyses in the *Corpus of Contemporary American English* (COCA) and the *News on the Web Corpus* (NOW) to substantiate the frequency and stability of specialized combining forms and their profitability in the formation of both novel and nonce words.

Keywords: neoclassical/abbreviated/secreted combining forms, ESP, word-formation

1. Introduction

Because of new scientific discoveries, medical advancements, and technological inventions, the lexicon of English is in continuous expansion (Arndt-Lappe et al., 2018). Many new words are recognized and repeatedly used by the speech community, therefore becoming part of English vocabulary as accepted neologisms. Others are coined for specific occasions and do not always become institutionalized or lexicalized (Brinton & Closs Traugott, 2005), yet their creation suggests how language may be enriched with new temporary vocabulary, namely nonce terms, connected to a historical, socio-political or economic period. The lexicon of English for Special Purposes (ESP), however, is rarely ephemeral, since the emergence of new specialized terms is generally motivated by lexical gaps which need to be filled.

The present paper analyses 30 combining forms retrieved from the *Oxford English Dictionary* (OED) and contributing to the lexical innovation of ESP. Combining forms (henceforth, CFs) are traditionally defined as neoclassical elements of Greek or Latin origin, such as *bio-* (Gr. *βίο*, Lat. *bio* 'life') or *-logy* (Gr. *λογία*, Lat. *logia* 'science'), which are added to other elements, either initially, as in *biodata* 'one's curriculum vitae', or finally, as in *musicology* 'the branch of knowledge that deals with music'. However, as recently demonstrated by Mattiello (2022), CFs comprise not only neoclassical bound elements, but also abbreviated forms, such as *e-* (from *electronic*) in *e-book*, *e-journal*, and secreted ones, such as *-nomics* (from *economics*, denoting 'the economic policies of a President or head of state') in *Clintonomics*, *Obamanomics* (cf. "classical" vs. "modern combining forms" in Prćić, 2005, 2008; Amiot & Dugas, 2021).

While in the literature neoclassical CFs are generally regarded as cases of compounding (Bauer, 2017, p. 150), abbreviated and secreted CFs have been attached different labels, including "affixoids" or "pseudo-affixes", which rather stress their derivational nature. Scholars also tend to confuse or merge CFs with the notion of "splinters" used in the formation of blends (Bauer, Beliaeva & Tarasova, 2019, p. 62), thus confirming their "fuzzy" nature and difficult categorization. In a recent study (Mattiello, 2022), the three categories of neoclassical, abbreviated, and secreted CFs have been considered as part of transitional morphology, i.e. intermediate between two subcomponents of word-formation: i.e., derivation and compounding. Like affixes, CFs are bound morphemes added to bases, yet like compound constituents, they have a high lexical density and can be stressed (e.g., *e-* has the same meaning as the adjective *electronic* from which it is clipped and is stressed in *é-mail*).

The dividing line between one category and the other is often not so well defined, nor are CFs classified in a univocal way in the linguistic literature. Section 2 tries to clear up the terminological confusion about CFs and related concepts. Section 3 explains the methodology followed for the selection and analysis of the data. Section 4 offers a qualitative scrutiny of the selected CFs and a quantitative analysis of the combining-form combinations that they produce, discriminating between highly frequent neological formations and occasionalistic nonce words. Finally, section 5 summarizes the research conducted and draws up some conclusions.

2. Combining forms: Definitions and Terminological Distinctions

Warren (1990, p. 112) was the first to stress that "combining forms are morphemes of a rather special kind", which represent a quite heterogeneous group of elements, differing in their origin as (1) allomorphs of model words (e.g., *astro-* in *astrodome*), (2) truncated forms of model words (e.g., *cyber-* in *cyberphobia*), or (3) parts of model words which happen to coincide with existing words (e.g., *-gate*, from *Watergate*, denoting 'an actual or alleged scandal' in *sexgate*). The same CF can even differ in the way it is interpreted depending on the word to which it is added: e.g., in *cheeseburger* 'hamburger topped with cheese', *-burger* is an abbreviated form, whereas in *fishburger* 'fried patty made of fish served in a bread-bun', it is secreted, in that a *fishburger* does not contain any beef meat. Moreover, CFs are divided into initial and final (*e-* vs. *-nomics*) depending on their position either to the left or to the right of a base. With regard to position, there are very few exceptions (i.e. *morph-/-morph* and *phil-/-phile*), which can occur both in initial (e.g., *morph-ology*, *phil-ology*) and final positions (e.g., *anthropo-morph*, *anglo-phile*) (Bauer, 1983; Plag, 2003). Initial CFs are close to prefixes while final CFs are close to suffixes.

This heterogeneity manifests itself in terminological confusion (e.g. "affixoids", "pseudo-affixes", "semi-affixes", "affix-like formatives", "folkmorphs", "splinters", along with "combining forms") and descriptive vagueness. Bauer already in the 1980s claimed that CFs belong to a type of word-formation which "has received very scant attention in the literature on morphology" (Bauer, 1983, p. 213). Scholars indeed tend to relegate this topic to the category of neoclassical CFs and only include discussions on neoclassical compounds in their description of English word-formation, disregarding completely the modern categories. More recently, Kastovsky (2009, p. 12) has even argued that the notion of CF is not necessary at all, as "[t]he categories of 'word', 'stem', 'affix', 'affixoid', 'clipping' and 'blending' necessary in word-formation for independent reasons are sufficient to deal with the formations in question".

Following recent studies on the topic (e.g., Amiot & Dugas, 2021), Mattiello (2022) has dealt with this phenomenon showing the importance of a finer-grained classification and an in-depth description of CFs. From her analysis, CFs consist of a miscellaneous set of morphological elements which share features such as semantic and phonological weight with compound constituents, but also features such as boundedness and fixed position (either before or after a base) with affixes. For these reasons, Mattiello (2022) accommodates the different categories of CFs under the same umbrella term of "transitional morphology", straddling the demarcation line between two subcomponents of word-formation, i.e. derivation and compounding.

In line with Mattiello (2022), CFs are here defined as initial or final bound morphemes which are either allomorphic variants of classical Latin or Greek words (e.g., *bio-* above), or shortenings of (native or non-native) English words (e.g., *econo-* from *economic* in *econo-politics*), often with the intervention of a secretion process (e.g., *-aholic* 'person addicted to' in *work-aholic* 'person addicted to work'). These are respectively called "neoclassical", "abbreviated", and "secreted CFs", and the complex words obtained from them are called "combining-form combinations".

A notion which is often confused with CFs is that of affixoids. The term "affixoid" is used within Construction Morphology by Booij & Hüning (2014), who note how the element *free* has developed the more general meaning 'without what is denoted by the base word' when it is used as the right constituent of compounds, as in *sugar-free* and *fat-free*. Moreover, unlike the suffix *-less* having the same denotative meaning in *sugar-less* and *fat-less*, the affixoid *-free* also presupposes a slightly negative evaluation (e.g., that the presence of sugar is not good for one's health). Therefore, affixoids exist alongside formally identical and usually free 'parent' morphs; yet, they acquire a more generalized meaning (Stevens, 2005, p. 73) or undergo a "desemanticization" process (Amiot & Dugas, 2021), as *-free* 'clear of something which or someone who is regarded as problematic' in *trouble-free*, *pollution-free*, *risk-free*, etc. Thus, like CFs, affixoids are "productive" (Bauer, 2001), i.e., are part of schematic constructions with high type frequency (e.g., [N-*free*]), and are bound to their bases (e.g., in *sugar-free*, the suffixoid and the base cannot be split up as **sugar very free*). However, unlike CFs, affixoids correspond to free-standing words, whereas in the case of CFs, correspondence to a free morpheme is accidental or synchronically irrelevant (cf. *-gate* 'political scandal' vs. the noun *gate*).

Still another crucial distinction is between CFs and blend splinters. Splinters are word parts that merge with one another or with full words in a blend (e.g., in *jeggings* ← *jeans* + *leggings* there are two splinters which graphically overlap, while in *Brexit* ← *British* + *exit* only the first element is shortened). Hence, the notion of splinter is very close to that of CF, differing from it only in their diverse degrees of productivity (Bauer, 2001). We may even envisage a diachronic evolution from blend splinter to abbreviated or secreted CF. Many splinters are indeed transitional from their status as blend constituents (i.e. word parts) to a more productive affix-like status as CFs (Correia Saavedra, 2013, 2016). This evolution is manifest in *-aholic*, which was originally attested in blends (e.g., *work-aholic* [1947], *choco-holic* [1961]), then became a productive CF included in the OED with the meaning 'a person addicted to –' (e.g., *carbo-holic* [1973], *shop-aholic* [1977], *news-aholic* [1979], *spend-aholic* [1982]), with a higher degree of frequency, productivity, and stability.

In this paper, the focus is on the CFs which have become productive and stable in ESP, because of their reiterated use in the formation of novel words by experts in specialized areas.

3. Dataset and Methodology

The methodology adopted for the selection of the CFs examined in this study combined an advanced search in the OED, using the designated tool, with a manual cleaning of irrelevant examples. The advanced search was conducted in November 2023. First, I selected the tab 'entries' of the OED and entered the search text 'combining form'. This search returned a list of 2,297 matching entries, which were then put in chronological order, by arranging results by 'date (newest first)'. Since my focus was primarily on recent forms, I added a temporal parameter to my search, by selecting the range 1950-today, and concentrated on this dataset. The 77 results obtained were the starting point for my analysis. To this list *robo-* was added manually because, even if its earliest attestation is before 1950, in 1988 it became a secreted form referring to something or someone 'resembling a robot, esp. in being resilient, emotionless, or futuristic'. Another CF that deserved inclusion, but did not match my advanced search because it is not dated in the OED, is *-nomics*. I decided to incorporate it because the earliest formation displaying this CF is dated 1969.

The final 79 results include all three categories – namely, neoclassical, abbreviated, and secreted forms – yet with some elements having evolved in time from abbreviated to secreted. This dataset was then manually reduced to 30 – 18 initial and 12 final – specialized CFs for the purposes of my analysis. Among the selected CFs, some are apparently similar in form, but actually diverge in meaning: e.g., *e-* is abbreviated from either *electric* (e.g., in *e-car*) or *electronic* (e.g., in *e-journal*). Hence it counts as two separate CFs (*e-^{1}* vs. *e-^{2}*) used in different specialized fields. The specialized fields of the selected CFs mainly involve digital and information technology, electronics, economics, and less relevantly chemistry, medicine, physiology, palΘontology, and linguistics.

In the corpus-based analysis, the three categories of neoclassical, abbreviated, and secreted CFs will be treated separately, also distinguishing between initial and final for each type, in order to verify their frequency and profitability in terms of new words obtained through this word-formation mechanism. Both type and token frequencies of occurrence in the *Corpus of Contemporary American English* (COCA) (Note 1) and in the *News on the Web Corpus* (NOW) (Note 2) will be shown, as well as their meanings and contexts of use. By observing quantitative data, a distinction between stable neologisms and nonce formations or hapax legomena (i.e. once-only attestations) will be also made. By exploring qualitative results, an evolution towards abbreviated and secreted CFs will be observed in ESP, as already remarked in general English (Mattiello, 2022).

4. Specialized Combining Forms in ESP: Stable and Unstable Formations

Specialized CFs are certainly expanding the lexicon of ESP with new words. The aim of this section is to establish the productivity of these morphological elements used in different areas and with diverse degrees of stability. Some of them are more stable and produce large sets of new words which deserve inclusion in lexicographic works, others, being more specific and technical, have a limited applicability, but still contribute to the lexical innovation of ESP.

4.1 Neoclassical Combining Forms

Neoclassical CFs mainly pertain to the areas of chemistry, biochemistry, and medicine, with rare exceptions.

Among initial CFs, *ichno-*, attested since 1956 and obtained from Greek ἴχνος 'track, trace', is used in palΘontology with reference to trace fossils, especially in terms relating to their taxonomy. In those terms, *ichno-* combines with either final CFs (e.g., *ichnography*, *ichnology*) or full words (e.g., *ichnofauna*, *ichnofossil*). A few of the new words are hapaxes (e.g., *ichnofamily* and *ichnographic*(*ally*)), yet most of them are stable neologisms (e.g., *This confirms the capacity of the body fossil record and **ichnology** to complement each other.* NOW, 17/03/2023).

Another specialized CF coming from Greek λεξικός 'lexicon' is initial *lexico-*, which has been attested from 1953 in modern linguistic terms denoting 'lexical and –'. It often combines with adjectives (e.g., *lexico-behavioural*, *lexico-grammatical*), but especially occurs before nouns (e.g. *lexico-dynamics*) or final CFs (e.g. *lexicography*). Nouns such as *lexicogrammar*, *lexicology*, *lexico-statistics* occur more frequently than the derivatives *lexico-grammatical*, *lexicologist*, *lexicostatistical*, *lexico-statistician*, with only one occurrence in either COCA or NOW. Nevertheless, many of these words (e.g., *lexicographer*, *lexicography*) are amply attested in dictionaries and corpora (e.g., *Davis studied at Rhodes University and Oxford before working in **lexicography** at the Oxford English Dictionary.* NOW, 07/01/12).

In chemistry, *seco-* (recorded since 1951) is a formative element used in naming derivatives, especially of steroids, in which fission of a ring has occurred, as in *seco-dicarboxylic* and *seco-steroid*. The origin of the CF is from Latin *secāre* 'to cut', followed by an *-o-* connective (Note 3). The use of this initial CF is very scanty, though, as the only word attested in corpora is *seco-steroid* (e.g., *The body metabolizes it into a **seco-steroid** hormone which, like*

all steroids, regulates genetic transcription, a basic and profound mechanism of action unlike any other vitamin. NOW, 14/06/18).

The CF *synapto-*, from Greek *συναπτ-ικός* 'connective', has been recorded in the OED from 1962. It is used as CF of 'synapse' in various terms in physiology: e.g., *synaptology*, *synaptosomal*, and *synaptosome*. Frequent formatives in corpora include *synaptogenesis*, *synaptonemal*, *synaptophysin*, *synaptotagmin*, whereas *synaptoclastic*, *synaptogenic*, *synaptologist*, *synaptoplasticity*, and *synaptotoxicity* occur only once either in COCA or in NOW (e.g., *Ongoing research supports psilocybin's ability to stimulate* **synaptogenesis** *and neuroplasticity in the brain.* NOW, 09/05/23)

Final CFs are more frequent in corpora than initial CFs in terms of both types and tokens. In chemistry and biochemistry, the final CF *-mer* (from ancient Gr. *-μερής* 'part, segment', first attested in 1975 in the OED) forms nouns denoting particular kinds of polymer (as in *dimer*, *oligomer*) or isomer (as in *epimer*, *tautomer*). It can also be suffixed to numerals (in full form or more usually as digits) to form nouns denoting polymers consisting of a given number of units (as *9-mer*, *16-mer*). Combinations with *-mer* mainly include stable words such as *copolymer*, *monomer*, *polymer*, *tetramer*, *trimer* (e.g., *This implied that the complex consisted of three identical subunits, arranged as a symmetrical* **trimer**. COCA, 01/08/18), with rare nonce words or hapaxes such as *25-mer* and *36-mer*.

Another frequent final CF is *-ogen*, first attested in 1961 in the OED and coming from an *-o-* connective and the Greek root *γίγνεσθαι* 'be born, become'. In biochemistry, it forms the names of an inactive precursor of a compound to whose name it is appended, as in *hypertensinogen*, *kininogen*, etc. English has inherited from French the names *hydrogène* and *oxygène*, which were adopted into English with the ending *-gene*, afterwards altered to *-gen*. On the analogy of these words, a considerable number of new terms have been added to the English vocabulary of chemistry, in which the ending -(*o*)*gen* expresses the sense 'that which produces'. They are usually names of chemical substances, as in *amidogen*, *cyanogen*, *nitrogen*, etc., or of classes of substances, as in *halogen*. The most frequent words in both corpora are *hydrogen*, *estrogen* (e.g., *The basic components of HRT are* **estrogen** *and progestin, which is a form of progesterone.* NOW, 10/11/23) and *nitrogen*, but novel nonce words such as *angiogen*, *astrogen*, *carbogen*, *entactogen*, *vitrogen* also demonstrate the profitability of *-ogen* in the coinage of new vocabulary.

The CF *-penia*, first attested in 1971 in the OED, originates from ancient Greek *πενία* 'poverty' and is primarily used in medicine. From the end of the nineteenth century, it has been found in a small number of English formations and adaptations of foreign words, apparently first in *leucopenia*, and later in *granulocytopenia* and *neutropenia*. It chiefly combines with first elements of Greek origin, as in *leukopenia/leucopenia*, *osteopenia*, *sarcopenia* (e.g., *Recent estimates suggest that* **sarcopenia** *affects 10% to 16% of the elderly population worldwide.* NOW, 13/10/23), and *thrombocytopenia*, which are stable formations, unlike the hapaxes *cytoglucopenia*, *erythroblastopenia*, *plancytopenia*, *reticulocytopenia*, and *vasopenia* attested only once in the corpora.

The CF *-valent*, from Latin *valentem* (present participle of *valēre* 'to be worth'), has been attested in the OED from 1977. It is a formative element occurring in a few words of general currency and in various scientific contexts, used with prefixes denoting number. For instance, in chemistry and immunology, it forms adjectives denoting 'having a valency of the specified number' (e.g., *trivalent*, *quadrivalent*, *tetravalent*). It is also used with Arabic numbers prefixed (e.g., *14-valent*). While *ambivalent* (with the prefix *ambi-*), *covalent*, *multivalent*, and *quadrivalent* are amply attested in the corpora (e.g., *A* **quadrivalent** *vaccine works by stimulating an immune response against four different antigens, such as four different viruses or other microorganisms.* NOW, 09/02/23), attestations of *hexacovalent*, *hypovalent*, *omnivalent*, and *plurivalent* are rare occurrences.

Table 1 summarizes the quantitative data resulting from the COCA and NOW corpora showing the productivity of specialized neoclassical CFs in terms of type frequency, token frequency, and hapax legomena.

Table 1. Frequencies of occurrence of neoclassical CFs in the COCA and NOW corpora

	TYPE FREQUENCY		TOKEN FREQUENCY		HAPAX LEGOMENA	
	COCA	NOW	COCA	NOW	COCA	NOW
ichno-	3	14	15	157	1	4
lexico-	14	12	188	1,803	3	1
seco-	1	1	3	1	0	1
synapto-	13	19	76	361	3	5
-mer	10	14	2,366	26,601	1	0
-ogen	38	83	24,198	145,067	5	16
-penia	14	29	380	3,135	2	5

-valent	20	43	2,990	13,529	3	7

The high frequency of final CFs such as initial *lexico-* and *synapto-* and final *-ogen* and *-valent* shows the "realised productivity" or "extent of use" (Baayen, 1993) of such morphological elements and their profitability (Corbin, 1987; Bauer, 2001, p. 49) for the expansion of ESP lexicon. The fact that these are also used to coin nonce terms attested only once in corpora (hapaxes) signals their "expanding productivity", measuring the rate at which these morphological categories are expanding and attracting new members. Some neoclassical CFs (e.g., *ichno-* and *seco-*) are quantitively less productive, but their lower frequency is correlated to a sector-related use and specificity.

4.2 Abbreviated Combining Forms

Abbreviated CFs do not have a classical origin, but are obtained by shortening existing words and combining them with bases. Many of them are used to coin new words in the areas of digital technology, electronics, and economics.

Among initial CFs, one of the most frequent is *cyber-*. Shortened from *cybernetic*, it was originally (from 1961) used to create words, mainly nouns relating to computers, information technology, and virtual reality, or denoting futuristic concepts. Later, *cyber-* was specifically used to form terms relating to the Internet, probably influenced by analogy to *cyberspace*. Formations include established terms, such as *cyberart, cyberattack, cybercommunity, cybercrime, cyberfriend, cyberlover, cyber-romance, cybersecurity, cyberspeak, cybersphere, cyberterrorism,* and *cyberworld*, as well as temporary nonce words (e.g., *cyber-family, cyber-debate, cyber-piracy, cyber-protest, cyber-secure*) whose number is daily increasing. The following example illustrates *cybercrime* in context: *In this age of rapid technological advancement and widespread internet usage, **cybercrime** has emerged as a grave menace to our society, economy and personal security.* (NOW, 23/11/23).

Shortened from the adjective *digital*, *digi-¹* is an initial CF attested from 1960 to form nouns denoting 'a digital device, system, etc.', as in *digibox, digicam*. Later its meaning has been reinterpreted and, as a result, the initial CF has become secreted (cf. § 4.3). Besides *digibox* and *digicam* (e.g., *Film high DR roll-off in the highlights is sorely missed with **digicams**.* NOW, 21/10/19), corpora attest the use of neologisms such as *digicode, digicorder, digipack, digiPad, digitech,* as well as nonce words occurring only once (e.g., *digicell, digi-drum, digiPod, digi-spoon,* and *digitape*).

As announced, the initial CFs *e-¹* and *e-²* are kept distinct because of their different origin and use. The initial CF *e-¹* is earliest attested in 1969 with the meaning *electric*, from which it is abbreviated. With this sense, it forms nouns denoting vehicles powered by electricity instead of or as well as more traditional means, as in *e-car, e-bicycle* (or *e-bike*), and *e-scooter*. In corpora, only stable nouns occur, such as *e-boat, e-bus, e-cycle, e-ferry, e-jeep, e-power, e-rail, e-street, e-train,* and *e-van*, although *e-car* and *e-scooter* recur more frequently (e.g., *The launching of the **e-scooter** follows the commencement of the construction of Electrum's electric scooter factory in June 2023.* NOW, 23/11/23).

As a separate entry, the OED offers another initial CF *e-²*, more recent [1988] and this time abbreviated from *electronic*. Found in numerous recognized formations from the late 1980s, as in *e-book* and *e-fit*, early formations are often preceded by corresponding compounds in *electronic* whose spelling is as two separate words (e.g. *electronic book* [1978] vs. *e-book* [1988], *electronic cash* [1967] vs. *e-cash* [1994]).

The CF *e-²* is indeed prefixed to nouns to indicate involvement in electronic media and telecommunications (esp. concerning the Internet), usually to distinguish objects or actions from their non-electronic counterparts. In the late 1990s and early 2000s, formations starting in *e-²* became very numerous and conspicuous in their frequency of use. The three specific semantic areas that these formations cover are: (1) terms relating to the publication or exchange of information in an electronic format (e.g., *e-edition, e-journal, e-publication, e-reader, e-text, e-zine*); (2) terms relating to computers and the Internet, i.e. computer-mediated and Internet-based (e.g., *e-fit* and other nonce words); and (3) terms relating to electronic financial transactions (e.g. *e-bill, e-cash, e-commerce, e-currency, e-dollar, e-money, e-ticket*). The latter more specialized meaning has been recorded from 1992. Besides *e-mail* (both noun and verb), the most common combinations are *e-book, e-cigarette, e-commerce* (e.g., *The business continues to deliver acceleration in **e-commerce** growth in the country.* NOW, 29/11/23), *e-learning, e-reader*, whereas nonce words in COCA include *e-product, e-progress, e-proof, e-quality, e-recruit, e-recycling, e-revenue, e-revival, e-reward, e-servicing, e-shopkeeper, e-smoker, e-smoking,* and *e-solution* among others.

The initial CF *econo-*, attested from 1964 in the OED, is found in a small number of formations. It is found earliest in *econometry/econometrics* 'the branch of statistical theory concerned with the analysis of economic phenomena' and related words, and in *econobox*, referring to 'an economy-sized box of something' or 'a nondescript economical car'.

As a CF, it can be shortened either from *economic* and prefixed to adjectives with the sense 'economic and –' (e.g., *econo-political*), or from *economy* and added to nominal bases. In the latter case, it can form nouns denoting

'things which are economical (or promote economy), inexpensive, or cheap' (e.g., *econocar*), or nouns denoting 'things or persons of a type involved in or associated with economics' (e.g., *econo-politics*). *Econometrics* and the related adjective *econometric* are the most frequent words in both corpora (e.g., *The popularity of **econometrics** shows the feelings of the people and entrepreneurs.* NOW, 16/09/23), yet several nonce words are also attested (e.g., *econocentric, econocratic, econography, economode, econopack, econoperformance, econo-size*, etc.).

The initial CF *lamino-* comes from the abbreviation of the adjective *laminal* 'produced by the blade of the tongue', with an additional *-o-*, and is used in the medical field. The first attested examples include *lamino-palatal* [1966] and *lamino-dental* [1968]. Other combinations found in corpora include *laminoplasty* (e.g., *Modified expansive open-door **laminoplasty** technique improved postoperative neck pain and cervical range of motion.* COCA, 01/03/19), *laminography, laminopathy, laminotomy*, and the nonce term *laminotomy*.

Abbreviated from *negative*, *nega-* was first attested in 1973. It is used to form the names of 'negative counterparts of things (chiefly units of measurement)' (e.g., *negawatt*), now especially in the context of energy (or other resources) saved as a result of conservation measures (e.g., *Basically, it said that RTOs, when organizing competitive auctions in wholesale markets, have to treat a megawatt of avoided energy use – a **negawatt** – the same way they treat a megawatt of energy.* NOW, 26/01/16). Examples of formatives attested in the COCA and NOW corpora include *negabarrel, nega-bobbing, negacaptain, negabot, nega-verse*, and the nonce words *nega-carbon, negacoal, negaholism, neganetwork, nega-oil, negatree, Nega-Trump, nega-cancer*, and *nega-input*.

The initial CF *petro-[1]*, shortened from *petroleum*, first occurred in the mid-twentieth century in *petrochemical* [1942] and *petrochemistry* [1942], although *petromax* (from G. *Petromax*) [1929] 'a proprietary name for a type of vaporised paraffin pressure lamp' and *Petro-Forge* [1969] 'a kind of forging machine powered by a petrol engine' also belong here. As an abbreviated form, it obtains terms relating to petroleum, especially with reference to the political and economic power of oil-producing countries, as in *petropolitics* (e.g., ***Petropolitics** come with financial risks as well as upsides.* NOW, 04/01/23), *petro-power, petro-resources*, and *petro-wealth*. Formations abound in corpora, including, among others, both *petrochemical, petrography, petrologist, petrology, petromania, petropolitics, petro-state*, and a conspicuous number of nonce words (e.g., *petro-domination, petro-fueled, petro-military, petro-nation, petro-oil, petro-plastic, petro-supplier*, etc.).

The CF *porta-*, shortened from the adjective *portable*, apparently appeared earliest in *portapak* [1951] 'a portable system comprising a video camera and recording equipment'. It is used to form the (often proprietary) names of various movable or portable versions of manufactured objects, devices, structures, etc., as specified by the second element: e.g., *portacam, portacrib, Porta-John* (from slang *john* 'a toilet, lavatory'), *porta-kit, porta-office, Porta-Phone, porta-printer, porta-screen*, etc. In corpora, we also find *portaball, portacabin* (e.g., ***Portacabin** schools were started by state government in 2012 to cater to tribals.* NOW, 19/01/23), *porta-cell, porta-chef, portal-gun, porta-potty*, and hapaxes such as *portabook, porta-box, portacomstyle, porta-filter, portaphone, portapool*, and *portaportal*.

The CF *syn-*, abbreviating *synthetic*, has been used from 1971 to form words denoting synthetic products, as in *syncrude* 'a synthetic product made from coal in imitation of crude oil', *synfuel, syngas, synjet, synoil*, and *synroc*. As such, it differs from the prefix *syn-*, a Latinized form of Greek *συν-* 'together' (e.g., *synergy, synthesis*), and generally carries stress. In corpora we mainly find stable formations, such as *syncrude, syndiesel, syn-free, syngas* (e.g., *Electrochemical splitting of CO_2 can create useful chemicals, such as **syngas**, Formic acid, Oxalic acid, and more.* NOW, 30/10/23), *synlawn, syn-turf*, and some rare nonce formations (e.g., *synsilk, synskin*).

The CF *-bot[1]* (from *robot*) is found in a number of formations from about 1966 onwards. The new words were originally blends involving *robot* (e.g., *mobot* from *mobile* and *robot*), but later *-bot[1]* was used to form nouns denoting 'a type of robot or automated device' or, in extended (metaphorical) use, to refer to 'a person regarded as an automaton', as in *fembot* [1976] 'a robot resembling a woman in appearance' and *nanobot* [1989] 'a nanorobot' (e.g., *The helical **nanobot** is made of silicon dioxide coated with iron.* NOW, 16/05/22) (cf. its specialized meaning in information technology, § 4.3). Other stable formations attested in corpora include *aerobot, autobot, chatbot, guardbot, medbot, microbot, minibot, newsbot, repairbot, sexbot, smartbot, teacherbot, twitterbot, waterbot*, whereas *agrobot, brainbot, girlbot, megabot*, and *securitybot* are some instances of nonce formations.

Attested in the OED from 1965, the abbreviated CF *-lect* (from *dialect*) is a terminal element used to designate a regional or social variety within a language, as in *idiolect* [1948] 'the linguistic system of one person, differing in some details from that of all other speakers of the same dialect or language'. It is also used in forming a number of technical terms in linguistics, as in *acrolect, basilect, isolect, sociolect*, etc. Hence, it has been converted to a noun and used to refer to 'a social variety of a language or dialect'. The most common formation in both corpora is *idiolect* (e.g., *Attackers can mimic the distinct **idiolect** of the target.* NOW, 18/11/23), followed by *Afrolect, basilect, ethnolect, interlect, multilect, sociolect*, while *hyperlect* and *topolect* are hapaxes.

The final CF *-olol* (from *propranolol*) is found in formations from the late 1960s onwards. It is typical of pharmacology, where it is used to form the names of 'beta-adrenoceptor blocking drugs derived from propranolol or having a similar molecular structure' (e.g., *acebutolol* 'a beta blocker which is partially selective for cardiac beta receptors, used in the treatment of hypertension, angina, and arrhythmia'). Specialized formations attested in corpora include *atenolol*, *bisabolol*, *bisoprolol*, *metabolol*, *metoprolol*, *nadolol* (e.g., *We aimed to test the effectiveness of isosorbide mononitrate as an adjunct to the b-blocker **nadolol** in the prophylaxis of first variceal bleeding in these patients.* COCA, 21/12/1996), *nebivolol*, *stanozolol*, *timolol*, while *penbutolol*, *pindolol*, *practolol*, *talinolol*, and *nandolol* are once-only occurrences.

In particle physics, *-onium* represents the abbreviation of *positronium*. Since 1987, it has been used to form the names of bound states of a particle and its antiparticle, as in *charmonium*, *nucleonium*, *toponium*. This is not to be overlapped with the homonymous antecedent form, occurring in chemistry from 1858 in *carbonium*, *hydrazonium*, *nitronium*, *phosphonium*, which instead comes from *ammonium* and forms the names of 'complex cations that contain a more or less electronegative central atom'. The only stable word attested in corpora is *charmonium* (e.g., *And, given that the specific quark is called "charm," its existence opens vast possibilities for puns – and that's without even getting into the fact that the technical term for the full family of particles containing these quarks is "**charmonium**".* NOW, 14/07/20), while *quarkonium* is a once-only formation attested in NOW.

The CF *-ylidene*, from *ethylidene*, earliest attested in 1971, is used in chemistry as an adaptation of a French chemical name. It is used to form the names of divalent radicals in which both valencies derive from the same atom, as in *benzylidene*. It replaces the *-idine* suffix when the name of the parent compound does not end in *-yl* (cf. *alkylidene*, *propylidene*, etc.). Corpora attest stable words such as *2-chlorobenzylidene*, *4-methylbenzylidene* (e.g., *An order came into force on Wednesday banning lotions containing oxybenzone, octinoxate, **4-methylbenzylidene** camphor or butylparaben from Thailand's marine national parks.* NOW, 04/08/21), *benzylidene*, *cyclopentylidene*, *nitrobenzylidene*, *o-chlorobenzylidene*, *retinylidene*, but *2-xhlorobenzylidene* is attested only once in NOW.

Table 2 summarizes the quantitative data resulting from the COCA and NOW corpora showing the productivity of specialized abbreviated CFs in terms of types, tokens, and hapaxes.

Table 2. Frequencies of occurrence of abbreviated CFs in the COCA and NOW corpora

	TYPE FREQUENCY		TOKEN FREQUENCY		HAPAX LEGOMENA	
	COCA	NOW	COCA	NOW	COCA	NOW
cyber-	423	282	8,821	8,687	129	64
digi-[1]	20	12	208	412	5	20
e-[1]	8	24	160	20,422	0	0
e-[2]	279	281	62,061	639,728	19	0
econo-	41	31	873	2,851	21	88
lamino-	2	6	31	70	1	0
nega-	15	11	32	123	8	5
petro-[1]	66	87	1,025	31,263	37	212
porta-	19	36	221	1,951	7	16
syn-	6	7	206	3,873	2	0
-bot[1]	35	43	285	10,055	9	166
-lect	4	10	33	96	1	1
-olol	13	20	145	628	3	3
-onium	1	2	6	9	0	1
-ylidene	2	9	4	36	0	2

As Table 2 shows, specialized abbreviated CFs are highly frequent in corpora, particularly initial *cyber-* and *e-[2]*, productively forming hundreds of types, but also initial *petro-[1]* and *econo-*, and final *-bot[1]*, used to form both neologisms and nonce words. The high number of tokens occurring in corpora and the numerous hapaxes obtained through these morphological elements confirm the stability of the CFs and their acceptability for novel formations and vocabulary expansion.

4.3 Secreted Combining Forms

Secreted CFs involve both abbreviation and reinterpretation. Formally, they are obtained from the back- or fore-clipping of a word. Semantically, they involve either a generalization or a specification process, which contributes to conferring a certain level of abstraction on the CF. Many of them are used to coin new words in the areas of digital technology and, less frequently, economics.

An initial CF which is not only abbreviated (see its concrete use for digital devices in § 4.2) but also secreted is *digi-²*. Attested from 1986 as a secreted CF, it is used to create 'nouns denoting a product, process, person, etc., relating to or characterized by the use of digital technology', such as *digi-age*, *digi-art*, and *digi-novel*. On some occasions, it is also prefixed to the names of styles of popular music (as in *digi-funk*, *digirock*, etc.) to form 'nouns denoting music incorporating sounds generated or modified digitally'. Hence, unlike *digi-¹*, *digi-²* does not refer to devices, but to more abstract entities such as art, age, or music, as in *digisphere* (e.g., *You can just launch and then let it disappear into the **digisphere**.* NOW, 02/06/19). Other stable formations attested in corpora include *digibook*, *digiconomics*, *digiconomy*, *digimarketing*, *digi-paint*, *digiphile*, *digishop*, *digitour*, *digizine*, while *digidance*, *digidata*, *digi-friendly*, *digimusician*, *digirama*, and *digitask* are nonce words.

The formation of the secreted CF *m-*, from *mobile*, is clearly analogical: its similarity with *e-²* (see § 4.2) is marked formally – both retain only the initial letter of a word, being rather borderline with initialisms – and semantically. Recorded earliest in *m-commerce* [1997], *m-* forms terms relating to 'commercial activity conducted through mobile electronic media and devices, especially mobile phones', as in *m-banking*, *m-payment*, *m-ticket*, etc.

Since 2001 it has also been used to create terms (occasionally temporary words and ad hoc formations) relating to 'social and cultural activity or phenomena conducted through or using mobile electronic media and devices', as in *m-government*, *m-health*, and *m-voting*, but earliest in *m-learning*. The analogical nature of these formations is confirmed by the existence of *e-²* parallel formations (cf. *e-commerce*, *e-learning*) (e.g., *Share of **m-commerce** into e-commerce is projected to reach 80 percent by 2024.* NOW, 29/11/23). Other recognized formations are attested in the COCA and NOW corpora: e.g., *m-cash*, *m-enterprise*, *m-finance*, *m-government*, *m-retailing*, *m-service*, *m-system*, *m-ticket*, while hapaxes in COCA include *m-banking*, *m-finance*, *m-healthcare*, *m-mailbox*, and *m-money*.

Besides being an abbreviated initial CF shortened from *petroleum* (e.g., *petrochemistry*, see § 4.2), *petro-²* is also secreted when it forms terms designating 'revenue, esp. foreign exchange, that derives from petroleum exports', as in *petrodollar* [1973] 'a notional monetary unit earned by a country from the export of petroleum', *petrocurrency* [1974] 'the currency of a petroleum-exporting country', *petrobillion*, *petro-naira*, and *petropound*. The secretion process activates semantic specification in this case. The most common word in corpora is *petrodollar* (e.g., *Since the menace of armed robbery went full throttle in the immediate **petrodollar** Nigeria of the early 1970s, it has grown further into becoming a social pandemic today.* NOW, 07/05/23), yet other formations are attested: e.g., *petrobank*, *petro-billionaire*, *petro-currency*, *petroeuro*, *petromoney*, *petroyuan*, and the nonce terms *petro-bourgeoisie*, *petro-charged*, *petro-shilling*, *petro-wealthy*, and *petro-welfare*.

The initial CF *robo-*, clipped from *robot*, has been attested from 1988 in a number of formations with the sense 'a/an – resembling a robot, especially in being resilient, emotionless, or futuristic'. Examples include both recognized nouns, such as *robocop* 'a robotic or bionic law enforcement officer' (used in science fiction from 1957) and *robocall* [1998] 'an automated telephone call which delivers a recorded message' (e.g., *But on Sunday Khatari cut a **robocall** to voters urging them to vote for Kagan.* NOW, 06/11/23), and nonce words such as *robo-boxer*, *robo-candidate*, and *robodisco*. In corpora, we also find *robo-adviser/robo-advisor*, *robocaller*, *robocar*, *robocat*, *robodoc*, *robodog*, *roboform*, *robohelp*, *robonaut*, *robo-signer*, *robosigning*, etc. Nonce words are surprisingly frequent, a few instances being *robobike*, *robobox*, *robobug*, *robobuilder*, *robodoctor*, *robohunter*, *robokiller*, *robomom*, and *robostaff*.

Shortened from the same base *robot*, the final CF *-bot²* is instead used in computing and information technology to form nouns denoting 'a type of automated program or (Internet) software, especially one which searches out information' (cf. also abbreviated *-bot¹* in § 4.2). More established combinations of this nature are *infobot* [1986] 'any of various automated systems for providing or obtaining information', *knowbot* [1988] 'a program designed to search through large numbers of databases and retrieve information', and *cancelbot* [1993] 'a program that searches for and deletes specified postings from Internet newsgroups'. Corpora also attest *Applebot*, *biobot*, *chatterbot*, *eventbot*, *filebot*, *Googlebot*, *junkbot*, *shopbot*, *spambot*, *talkbot*, *Twitterbot*, etc.

Some of the less common or more ephemeral forms are *searchbot*, *spybot* (e.g., *Finish them, and grab the **Spybot** before sliding down to the bazaar.* NOW, 10/06/21), and *warbot*. A semantic specification process has intervened in the formation of this CF.

The final CF *-nomics* is found in formations from the first half of the nineteenth century in the sense denoting the science or study of a subject specified by the first element, as *pyronomics*, *phoronomics*, etc. However, from the second half of the twentieth century, it has formed nouns denoting (often semi-humorous) fields of *economics*, from which it is shortened, as specified by the first element. It is therefore reinterpreted, via a specialization process, as 'the economic policies of –', the second element often being the name of a president of the United States, as in *Nixonomics* [1969] (from Richard *Nixon*), *Reaganomics* [1970] (from Ronald *Reagan*), and *Clintonomics* [1992] (from Bill *Clinton*), or another famous leader or political authority, as in *Rogernomics* [1985] (from the name of *Roger* Owen Douglas, New Zealand Minister of Finance).

Other attested formations include *Abenomics*, *Berlusconomics*, *Bushonomics*, *Carternomics*, *Gandhinomics*, *Obamanomics*, *Osbornomics*, *Popenomics*, *Putinomics*, *Thatchernomics*, *Trumponomics* (Note 3), and the nonce terms *corruptonomics*, *Dixonomics*, *dole-nomics*, *Dollarnomics*, *Gorbanomics*, *Wikinomics*, *womanomics*, some of which extend the base to a common noun (i.e. *corruption*, *dollar*, *woman*). A very recent formation occurring 2,186 times in NOW is *Bidenomics* (e.g., *U.S. President Joe Biden touts the administration's "**Bidenomics**" agenda during an October speech... from the White House.* NOW, 22/11/23).

The final CF -*verse* (from *universe*) has been used from 1981 to form nouns denoting 'the sphere or realm of what is specified or indicated by the first element'. Before becoming a recognized CF, the blends *nulliverse* [1847] 'a world devoid of any unifying principle or plan' and *multiverse* [1895] 'the universe considered as lacking order or a single ruling and guiding power' occurred. Nowadays, -*verse* is frequently used in the context of computing and electronic communications, as in *blogiverse*, *digiverse*, and *Twitterverse* (e.g., *The **Twitterverse** further said that the airline just followed the rules.* NOW, 12/08/23). Other formations are attested in corpora: *Appleverse*, *megaverse*, *metaverse*, *negaverse*, *pluriverse*, *televerse*, *webverse*, etc. However, from 1993, it has also obtained nouns denoting 'the fictional world associated with a specified character, television series, author, etc.', as in *Potterverse* (from J. K. Rowling's *Harry Potter* series), *Honorverse* (from David Weber's *Honor Harrington* series), *Buffyverse* (from the black comedy fantasy film and later television series *Buffy the Vampire Slayer*), *Whoniverse* (from the television series *Doctor Who*), and the nonce terms *Bronteverse*, *Garfieldiverse*, *Marvelverse*. Therefore, the CF acquires more specific meaning in the secretion process.

Table 3 summarizes the quantitative data resulting from the COCA and NOW corpora showing the productivity of specialized secreted CFs. The absence of hapax legomena in NOW is generally related to the fact that the automatic search is limited to 1,000 results per CF.

Table 3. Frequencies of occurrence of secreted CFs in the COCA and NOW corpora

	TYPE FREQUENCY		TOKEN FREQUENCY		HAPAX LEGOMENA	
	COCA	NOW	COCA	NOW	COCA	NOW
digi-[2]	12	21	26	519	6	0
m-	9	23	38	2,643	5	0
petro-[2]	7	21	235	2,331	3	2
robo-	81	128	1,147	19,567	39	19
-bot[2]	7	12	117	551	2	0
-nomics	10	43	339	4,155	3	9
-verse	16	54	828	12,856	6	3

Table 3 indicates that secreted CFs are less frequent than neoclassical and abbreviated CFs, in that they represent a rather recent phenomenon, still ongoing and sometimes developing from blends or abbreviated forms. In most cases, indeed, the shortening process precedes the semantic reinterpretation process, as with *digi-[2]*, *petro-[2]*, and -*bot[2]*, all corresponding to antecedent abbreviated forms (see § 4.2). The scarce number of hapaxes also goes in the same direction, confirming that secreted CFs are contributing to the development of ESP less conspicuously than neoclassical and abbreviated forms. Nevertheless, some less specific CF, like initial *robo-*, is producing both types and hapaxes, showing an increasing productivity and going towards stability and recognition.

5. Conclusions

The present study has investigated the role of combining forms in the development of ESP lexicon. Although all the CFs examined here are recorded in the OED, their productivity is increasing and the stability of the combining-form combinations resulting from them has been confirmed by corpus linguistic analysis. Corpus-based results have demonstrated that, besides neoclassical CFs already attested in ESP since the nineteenth century, abbreviated and secreted CFs are also currently contributing to the lexical expansion of specialized English.

Some of the CFs analysed have formed series of new words, especially stable and recognized neologisms, but also nonce words or hapax legomena which are attested only once in corpora, thus showing the potential profitability of these morphological elements in the coinage of still novel formations.

CFs are transitional phenomena often originated from a blending process, or from a shortening process followed by semantic reinterpretation. Secreted CFs, which involve semantic specialization or generalization, represent the last step of the morphological evolution from blend splinters to productive morphological elements that are very close to affixes. Hence, they are less frequently found in corpora, but their expanding productivity should be monitored in order to show how relevant they are to the vocabulary expansion of ESP.

The cases of *net-* (from *Internet*) and *-flation* (from *inflation*) belong here, in that they are splinters originated from blends (e.g., *netiquette* ← *Internet* + *etiquette*; *stagflation* ← *stagnant* + *inflation*), but they are becoming productive in the formation of novel words, such as *slumpflation* and *netizen* or *Netscape* (Note 5). Thus, the two splinters are potential CFs to be recorded in dictionaries not as part of blends, but as separate entries (i.e. initial CF *net-* 'related to Internet' and final CF *-flation* 'increasing inflation accompanying the state of the economy'), in spite of the scarce number of novel formations found in corpora.

Hapaxes in corpora, indeed, are often attested only once because of their sector-relatedness and specificity, but tend to be recognized very soon by the community of experts who are familiar with the lexicon of ESP and to be recurrently re-used in professional contexts as a confirmation of their lexicalization and stability.

References

Amiot, D. & Dugas, E. (2021). Combining forms and affixoids in morphology. In R. Lieber (Ed.), *The Oxford Encyclopedia of Morphology*. Oxford, Oxford University Press. https://www.oxfordreference.com/view/10.1093/acref/9780190682361.001.0001/acref-9780190682361-e-569

Arndt-Lappe, S., Braun, A., Moulin, C., & Winter-Froemel, E. (Eds.) (2018). *Expanding the Lexicon: Linguistic Innovation, Morphological Productivity, and Ludicity*. Berlin/Boston, De Gruyter.

Baayen, H. R. (1993). On frequency, transparency and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1992* (pp. 181–208). Dordrecht, Kluwer Academic Publishers.

Bauer, L. (1983). *English Word-formation*. Cambridge, Cambridge University Press.

Bauer, L. (2001). *Morphological Productivity*. Cambridge, Cambridge University Press.

Bauer, L. (2017). *Compounds and Compounding*. Cambridge, Cambridge University Press.

Bauer, L., Beliaeva, N. & Tarasova, E. (2019). Recalibrating productivity: Factors involved. *Zeitschrift Für Wortbildung / Journal of Word Formation*, *3*(1), 44–81.

Booij, G. & Hüning, M. (2014). Affixoids and constructional idioms. In R. Boogaart, T. Colleman & G. Rutten (Eds.), *Extending the Scope of Construction-based Grammar* (pp. 77-106). Berlin/Boston, De Gruyter.

Brinton, L. J., & Closs, T. E. (2005). *Lexicalization and Language Change*. Cambridge University Press. https://doi.org/10.1017/CBO9780511615962

Corbin, D. (1987). *Morphologie dérivationnelle et structuration du lexique*, 2 Vols. Tübingen, Max Niemeyer Verlag.

Correia Saavedra, D. (2013). *Lexical Blending in English: From Splinters to Combining Forms*. Unpublished Master Thesis. Faculty of Humanities, University of Lausanne.

Correia Saavedra, D. (2016). Automatically identifying blend splinters that are morpheme candidates. *Digital Scholarship in the Humanities*, *31*(1), 55–71.

Kastovsky, D. (2009). *Astronaut*, *astrology*, *astrophysics*: About combining forms, classical compounds and affixoids. In R. W. McConchie, A. Honkapohja & J. Tyrkkö (Eds.), *Selected Proceedings of the 2008 Symposium on New Approaches in English Historical Lexis (HEL-LEX 2)* (pp. 1–13). Somerville, MA: Cascadilla Proceedings Project.

Mattiello, E. (2022). *Transitional Morphology. Combining Forms in Modern English*. Cambridge, Cambridge University Press.

Oxford English Dictionary Online (1989–). 2nd/3rd Edn. Oxford, Oxford University Press. http://www.oed.com/.

Plag, I. (2003). *Word-formation in English*. Cambridge, Cambridge University Press.

Prćić, T. (2005). Prefixes vs initial combining forms in English: A lexicographic perspective. *International Journal of Lexicography*, *18*(3), 313–334.

Prćić, T. (2008). Suffixes vs final combining forms in English: A lexicographic perspective. *International Journal of Lexicography*, *21*(1), 1–22.

Stevens, C. M. (2005). Revisiting the affixoid debate. On the grammaticalization of the word. In T. Leuschner, T. Mortelmans & S. de Groodt (Eds.), *Grammatikalisierung im Deutschen* (pp. 71–83). Berlin/New York, De Gruyter.

Warren, B. (1990). The importance of combining forms. In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer & J. R. Rennison (Eds.), *Contemporary Morphology* (pp. 111–132). Belin/New York, De Gruyter.

Notes

Note 1. The *Corpus of Contemporary American English* contains more than one billion words (more than 25 million words each year 1990-2019) from eight genres including spoken language, fiction, popular magazines, newspapers, academic texts, TV and film subtitles, blogs, and other web pages. https://www.english-corpora.org/coca/

Note 2. The *News on the Web Corpus* contains 18.3 billion words of data from web-based newspapers and magazines from 2010 to the present time. More importantly, the corpus is constantly updated and grows by about two billion words each year. https://www.english-corpora.org/now/

Note 3. See Plag (2003, pp. 155-159) for the medial *-o-* that often appears in "neoclassical compounds".

Note 4. Occurring with different spellings, namely, *Trump(-)onomics*, *Trumpenomics*, *Trumpnomics*, and *Trumpanomics*.

Note 5. The OED records the following blends: *netiquette* [1982] 'an informal code of practice regulating the behaviour of Internet users when using email, bulletin boards, chat rooms, newsgroups, etc.', *netizen* [1984] 'a person who uses the Internet, esp. habitually', *Netscape* [1988] 'a proprietary name for: a browser used to access and display documents on the World Wide Web and other computer networks'; *stagflation* [1965] 'a state of the economy in which stagnant demand is accompanied by severe inflation', *slumpflation* [1974] 'a state of economic depression in which decreasing output and employment in industry are accompanied by increasing inflation'.

Copyrights

The Challenges of Creating a Bilingual (English-Italian) E-Dictionary of Sports and Games Terminology

Daniele Franceschi[1]

[1] Department of Foreign Languages, Literatures and Cultures, 'Roma Tre' University, Rome, Italy

Correspondence: Daniele Franceschi, Department of Foreign Languages, Literatures and Cultures, 'Roma Tre' University, Via Ostiense 234-236, 00146 Rome, Italy. E-mail: daniele.franceschi@uniroma3.it

Abstract

The aim of this paper is twofold. It firstly intends to provide an overview of the existing bilingual English-Italian dictionaries of sports and games terminology, in order to illustrate their main features and consider how this area of ESP (English for Special Purposes) lexicography may be improved. In the second part, a hypothetical e-dictionary entry is presented, taking as an example the lemma *padel*, indicating a new discipline that has become very popular in the past few years. The advantages of using online multimodal and multimedial dictionaries are discussed, while also making reference to the challenges of creating them: on the one hand, they make it possible to incorporate a plethora of information also via external hyperlinks, which makes them more dynamic resources with an encyclopedic character, but on the other hand it is necessary to constantly update dictionaries as hyperlinks 'age' over time and new disciplines appear. The inclusion of authentic material may also be problematic due to copyright reasons.

Keywords: bilingual lexicography, English-Italian, sports, games, terminology

1. Introduction

Sports and games are pervasive in our society both as competitive and as recreational activities. In some way or another, they affect us all, either because we engage in certain disciplines, professionally or amateurly, or simply because we are exposed to mega-sport events such as the Football World Cup, the Olympic Games and a number of other minor contests. Regardless of whether we are active participants or passive spectators, we all need, in different degrees, to use or at least understand the language of sports and games as it appears in the press, on social media and on TV.

Due to the presence of terminology, this may not always be an easy feat. In addition, as new disciplines emerge and grow (some examples are bouldering, padel, parkour, to mention just a few), languages constantly incorporate new words. English is a generous 'donor' of sports terminology to other languages. This is due, on the one hand, to its established lingua franca role also in sports events, which bring together athletes, teams and coaches from different countries. On the other hand, it is also true that many sports and games originate(d) in English-speaking environments before being exported to other parts of the world (Note 1). Hence, the abundance of anglicisms related to sports and games in the world's languages.

In the case of Italian, the introduction of English loanwords is also facilitated by an "open attitude of Italian society to Anglo-American culture, and the absence of any purist language policy as far as lexical borrowing is concerned" (Pulcini, 2008, p. 140). This has produced the co-existence of certain anglicisms with their Italian counterparts (e.g., En. *coach* and It. *allenatore*, En. *corner* and It. *calcio d'angolo*, etc.), but also the input of terminology which is now fully integrated in Italian (e.g., *derby* in football, *K.O.* in boxing; the names of various disciplines, such as *kickboxing*, *bungee jumping*, *standup paddleboarding*, and so forth) (Note 2).

Despite the pervasiveness of the language of sports and games, which may also include jargon and technical terms, dictionaries in this field do not abound or they tend to give prominence to the most popular sports. The 2000s, for instance, have seen the appearance of several dictionaries of football in multilingual editions (Pons, 2006; Yıldırım, 2006; Taborek, 2020), while the ones covering other disciplines are still few; some of them were published on the occasion of important events, such as the Olympic Games (Wehlen, 1976), the multi-sport European Championships (Sirges, 1980) or the various World Cups (Binder & Brasse, 1998).

This paper examines the existing bilingual English-Italian dictionaries of sports and games terminology and discusses the challenges of updating them, not just in terms of their contents but also in their form. It is indeed an undeniable fact that lexicographic resources need to be more practical in today's mobile society (Fuertes Olivera, 2018; Jackson, 2018) and be transformed into multimodal digital products in order to be more easily accessible and satisfy the modern user's needs. Section 2 illustrates the features of the two main English-Italian dictionaries in this area, namely Bonanno's (1988) *Dizionario dello Sport e di Medicina Sportiva* and Ragazzini's (1998)

Dizionario dello Sport, but also examines some sections of *Cambridge Word Routes – Dizionario Tematico dell'Inglese Contemporaneo* (1995), which specifically address certain sports disciplines. Section 3 provides an overview of minor online lexicographic resources, while section 4 proposes an improved English-Italian e-dictionary entry and exemplifies the sort of features that modern online lexicographic resources should have in order to be of practical use among translators, language learners and anyone interested in the micro-language of sports and games in English and Italian for professional, business or more personal purposes. Section 5 concludes the paper with some final considerations and a brief discussion of the future of lexicography.

2. Bilingual English-Italian dictionaries of sports and games terminology

The 2000s have not seen the publication of the revised editions of the two main bilingual English-Italian dictionaries mentioned above, which have remained in paper form and have not even been made available on CD-ROM. In addition, Ragazzini (1998) is now out of print and no longer easily found. Bonanno (1988) is, instead, still published today despite its being an older dictionary. Similarly, *Cambridge Word Routes – Dizionario Tematico dell'Inglese Contemporaneo* (1995) has not been updated, is not in electronic format, but is still available.

2.1 Dizionario dello Sport e di Medicina Sportiva (Bonanno, 1988)

This dictionary was compiled by an Italian secondary school teacher of English with the aim of contributing "to the understanding of English or American terms of many sports, even those not well known or practiced in our country" (Note 3). Therefore, it mainly targets Italian speakers, even though it is not unidirectional because it contains two sections with word lists and translation equivalents in both Italian and English. The merit of the work, which covers 63 disciplines and includes over 20,000 terms, lies in the fact that it always endeavors to provide the Italian counterpart or at least an explanation of the English terms even when they have made their way into Italian (e.g., *skateboard* 'monopattino'). The aim is to provide an incentive not to resort too frequently and casually to anglicisms that may be unclear to those unfamiliar with the English language and which refer to concepts that can be perfectly expressed in Italian too. An attempt has been made to list the actual terms used by specialists in the various disciplines: the author gathered materials from different sports federations, such as C.O.N.I. (Italian National Olympic Committee), as well as from the British Council and the cultural adviser to the Quebec delegation in Italy. The dictionary, however, is not based on a real corpus, but just informed by a variety of authentic data. Some lemmas are accompanied by black and white drawings that appear either on the right or left margin of the page or in the body of the entries. It is not clear, though, what the rationale behind the use of illustrations is. While some drawings are indeed justified because they describe complex notions, such as certain procedures and techniques, e.g. the five-step delivery mode in bowling, others simply represent objects, e.g. a baseball bat, a baseball ball, etc., and are not really necessary for comprehension. They only seem to have a sort of decorative function.

The main shortcoming of the dictionary is that the lemmas do not come with proper definitions, phonetic transcriptions or example sentences, but simply with their equivalents in English or Italian (Table 1), which are not of much help if one needs more than just a translation. Since terms are not presented in their context of use, the work looks more like a glossary than a real dictionary.

Table 1. Example of dictionary entry in Bonanno (1988)

pall-mall, s., pallamaglio

pan, s., (*vela*), aggottatoio

Knowing that *pall-mall* corresponds to 'pallamaglio' in Italian does not really explain what the discipline consists of (Note 4). Similarly, neither *pan* nor 'aggottatoio' in the language of sailing are transparent unless one is already familiar with one of the two terms and is just looking for its counterpart in the other language (Note 5). This is where the use of an illustration would have been particularly useful. In other cases, not only are proper definitions missing, but the explanations provided are also somewhat vague and inaccurate: *paddle tennis*, for instance, is only generically defined as 'a kind of tennis' ('specie di tennis'), thus failing to distinguish it from other racket sports which also resemble tennis, e.g. squash, ping-pong or badminton. In addition, certain lemmas, which sometimes consist of phrases and fixed expressions (Note 6) are not faithfully reproduced in Italian. Consider, for instance, *score (to) the winning goal in the last minute*, which could have been literally translated as 'segnare il gol vincente all'ultimo minuto'; the dictionary instead opts for 'segnare il gol vincente in zona Cesarini', which is both culture-bound and stylistically marked. Albeit stable and lexicalized, the Italian expression is very colloquial. It originated as a reference to an Italian-Argentinian footballer, Renato Cesarini, who often scored goals in the last few minutes of a match. Therefore, doing something in 'Cesarini zone', in a wider sense, means that whatever we refer to is happening at the very last moment, before it is too late.

The dictionary never provides any information about style or register, thus failing to address the differences in connotation between the English and Italian terms. This is true not just cross-linguistically, but also within the same language: a case in point are near-synonyms, which are distinguished neither in terms of their subcategorization frame, i.e. by making reference to their syntactic behavior, nor stylistically. The Italian word 'applauso', for instance, is made to correspond to both *applause* and *cheer*, but without any examples contrasting them it is impossible for dictionary users to understand whether and to what extent the two English words can be interchangeable.

Bonanno (1988) also presents some ambiguities in the indication of the sport provided in brackets after each term. It is not clear, for instance, what discipline the abbreviated form 'tiro' refers to, because it could be either archery ('tiro con l'arco') or shooting ('tiro a segno/volo'). This has implications for the comprehension of certain lemmas. If we consider *back of blade*, which is associated to 'tiro', it cannot be determined what it stands for; its Italian counterpart ('costola') too is somewhat vague due to its polysemic nature. Most probably, *back of blade* indicates the part of the bow that faces the archer when the bow is held in shooting position, but it may also indicate a part of the weapon used in sabre, a discipline of modern fencing. It is also interesting to observe that the dictionary omits the definite article (*back of the blade*), thus causing a potential problem for the correct use of the term.

Lastly, there appears to be a gender bias in the representation of certain disciplines. Gymnastics tends to be associated to women: while *women's gymnastics* is present as a separate entry, *men's gymnastics* is not; players are viewed as male by default and the marked form *lady player* is proposed for Italian 'giocatrice'; the words *champion* and *winner* only appear with their masculine counterpart in Italian ('vincitore'), but never in the feminine form ('vincitrice'), which is instead proposed as a translation for *defeater*.

2.2 Dizionario dello Sport (Ragazzini, 1998)

It can certainly be argued that, compared to Bonanno (1988), this dictionary introduces several improvements. To begin with, it is more comprehensive because it consists of approximately 50,000 entries covering the most widely practiced sports in Italy and in English-speaking countries (from soccer to baseball, basketball, cricket, field hockey, rugby, tennis, golf, water and winter sports, athletics, horseback riding and equestrianism, motoring, cycling, bowling, combat sports, billiards and many others). In addition, the first part, the English-Italian section, is typically organized as a proper dictionary with the inclusion of nested sub-headwords instead of listing all entries alphabetically as separate lemmas. Tables 2 and 3 show the entry for *score*, both as a noun and as a verb, in Bonanno (1988) and Ragazzini (1998), respectively. What is immediately clear is that the latter is much more detailed (even without including *score again*, *score against*, *score off*, *score on* and *score up*, which are listed separately) and that the arrangement of the entry follows the standards of modern lexicography. Therefore, it does not present itself as a mere glossary of terms as in the case of Bonanno (1988).

Table 2. Entry for *score* in Bonanno (1988)

score, s., punteggio
score, s., risultato
score, s., segnatura
score (to), v. t., segnare
score (to), v. i., segnare il punteggio
score (to), v. i., (*calcio*), far centro
score (to) a goal, (*pallc.*), segnare un canestro
score (to) a goal, (*calcio*), segnare un gol
score (to) a hit, fare un punto
score (to) an advantage, ottenere un vantaggio
score (to) a point, segnare un punto
score (to) a success, riportare un successo
score (to) a try, (*rugby*), segnare una meta

Table 3. Entry for *score* in Ragazzini (1998)

score, n. *1* punteggio; risultato: **the final s.**, il risultato finale *2* (*calcio, ecc.*) segnatura, marcatura (*il singolo punto realizzato*); gol; rete *3* (*cricket*) 'score': (*di una squadra*) **to make a s. of a certain number of points**, fare (*o conseguire*) uno score di un certo numero di punti *4* (*volley*) 'score'; punteggio. • (*calcio, ecc.*) a s. draw, un pareggio con segnatura di gol (*1-1, 2-2, ecc.*) • (*calcio*) **a s. like that of a tennis set**, un punteggio tennistico (*6-1, 5-0, ecc.*) • (*calcio, ecc.*) **a s. of four goal**, una quaterna (*4 gol segnati*).
to **score**, A *v.t. 1* aggiudicare; assegnare, attribuire (*punti, ecc.*): (*tennis*) **to s. a game for a player**, aggiudicare un game a un giocatore; (*boxe*) **The German judge scored him 20** (*o* **scored 20 to him**), il

giudice tedesco gli attribuì 20 punti *2* (*boxe*) mettere a segno, portare (*un colpo*): **He scored a direct hit with his first shot**, al primo colpo mise a segno un diretto *3* (*calcio, ecc.*) segnare, fare, realizzare, mettere a segno, siglare (*un gol, un canestro, un punto*): **Jack has scored four goals**, Jack ha fatto quattro gol; **They've scored three goals against us in half an hour**, ci hanno fatto tre gol (*o* il nostro portiere ha incassato) in mezz'ora; (*tennis*) **to s. a lot of points with one's serve**, fare un mucchio di punti con il servizio *4* (*scherma*) mettere a segno: **The winner of a bout is the fencer who first scores five hits**, il vincitore di un assalto è lo schermitore che per primo mette a segno cinque stoccate. **B** *v.i. 1* (*basket, calcio, ecc.*) fare punti; andare a segno; segnare; realizzare; fare centro; (*basket*) fare canestro; andare a canestro; centrare il canestro; (*calcio, ecc.*) fare gol; fare rete; andare in gol; andare a rete; marcare; centrare la porta; insaccare; finalizzare: **to s. early**, segnare all'inizio (*della partita*); **to s. regularly**, segnare con regolarità; (*calcio*) **to s. for the Glasgow Rangers**, segnare per i Glasgow Rangers; (*basket*) **to s. from the keyhole**, centrare dalla lunetta; **We scored in the second half**, segnammo nel secondo tempo; la nostra squadra passò nel secondo tempo *2* tenere il punteggio; fare il segnapunti. • (*basket, calcio, ecc.*) **to s. after a fast break**, segnare in contropiede; castigare gli avversari in contropiede • (*calcio, ecc.*) **to s. at the far post**, insaccare sul secondo palo • **to s. at will**, segnare quando si vuole (*o* a proprio piacimento) • (*basket*) **to s. a basket**, fare canestro, fare cesto; andare a canestro; centrare il canestro; mettere dentro (*fam.*) • (*basket*) **to s. a basket after a fast break**, fare canestro in contropiede (*basket*) **to s. a basket with a bank shot**, fare canestro su tiro di tabellone • (*calcio*) **to s. by heading the ball**, segnare (*o* fare gol) di testa • (*cricket*) **to s. a century**, segnare (*o* realizzare) una centuria • (*calcio, ecc.*) **to s. a chance**, realizzare (*o* sfruttare) un'occasione da gol • (*calcio, ecc.*) **to s. a the consolation goal**, fare (*o* segnare, cogliere) il gol della bandiera • (*calcio*) **to s. a curling free kick**, segnare con un calcio di punizione a spiovere • (*calcio*) **to s. a direct hit with one's first shot**, centrare la porta al primo colpo (*o* di prima intenzione) • (*calcio, ecc.*) **to s. easily on the rebound**, non avere difficoltà a segnare sul rimbalzo • (*calcio, ecc.*) **to s. an equal number**, segnare lo stesso numero di gol • (*calcio, ecc.*) **to s. the equalizer** (*o* **the equalizing goal**), segnare il gol (*o* il punto) del pareggio; recuperare un gol; pareggiare • (*calcio, ecc.*) **to s. first** (*o* **the first goal**), essere i primi a segnare; aprire le marcature • (*calcio, ecc.*) **to s. freely**, segnare agevolmente • (*calcio, ecc.*) **to s. freely after a counterattack**, castigare gli avversari in contropiede • (*calcio*) **to s. from the corner**, segnare (*o* realizzare) su calcio d'angolo • (*calcio*) **to s. from a penalty** (**kick**), segnare su rigore; trasformare un rigore • (*calcio*) **to s. from a penalty kick the first time one takes it [the second round]**, trasformare un rigore in prima battuta [in seconda battuta] • (*calcio*) **to s. from the penalty spot**, trasformare dal dischetto • **to s. a goal**, (*basket, netball*) fare un canestro; fare cesto; centrare il canestro; andare a canestro; (*calcio, hockey, pallanuoto, ecc.*) fare un gol; segnare un gol; fare una rete; realizzare un gol; andare a rete; centrare la porta; mettere dentro (*fam.*) • (*calcio*) **to s. a goal after a fast break**, fare un gol in contropiede • (*rugby*) **to s. a goal by a drop kick**, trasformare un calcio di rimbalzo (*vale tre punti*) • (*calcio*) **to s. a goal during a scuffle**, segnare in mischia • (*calcio*) **to s. goals from set pieces**, segnare da fermo (*su corner, su punizione, su rigore, ecc.*) • (*calcio, ecc.*) **to s. a goal to a yet unbeaten goalkeeper**, violare la rete avversaria • (*calcio*) **to s. a hat kick**, fare una tripletta • (*calcio*) **to s. a header**, fare un gol di testa • **to s. a hit**, (*calcio, ecc.*), andare a segno, segnare; (*boxe*) mettere a segno un colpo; (*scherma*) mettere a segno una stoccata • (*calcio, ecc.*) **to s. in the first minutes** (*o* **moments**) **of play**, segnare nei primi minuti di gioco; colpire a freddo • (*calcio, ecc.*) **to s. in the last minute**, segnare all'ultimo momento • (*calcio, ecc.*) **to s. into an unguarded net**, segnare a porta vuota (*o* a portiere battuto) • (*calcio, ecc.*) **to s. a lead** (*o* **the leading goal**), andare in vantaggio; portare in vantaggio la propria squadra • **to s. a run** (*baseball*), realizzare un 'run'; fare un punto con un 'home run' (*q.v.*); segnare un punto facendo il giro del campo fino alla casa base; (*cricket*) fare un punto con un 'run': **In cricket a run is scored whenever the two batsmen exchange their wicket positions on hit or passed balls without either being out**, nel cricket si segna un 'run' ogni volta che i due battitori fanno la corsa per il cambio di wicket dopo che la palla è stata battuta ovvero è passata, purché nessuno dei due si trovi fuori dalla sua zona (*di norma un punto va a entrambi*) • (*calcio, ecc.*) **to s. the** (*o* **one's**) **second goal** (*o* **to s. one's second**), raddoppiare: **England have scored their second**, l'Inghilterra ha raddoppiato • (*nelle corse*) **to s. a solo win**, arrivare primo (al traguardo) da solo • **to s. a success**, ottenere (*o* riportare) un successo • (*calcio, ecc.*) **to s. a teammate**, mettere un compagno in condizione di segnare • (*rugby*) **to s. a try**, segnare (*o* realizzare) una meta; andare in meta (*facendo 4 punti*) • (*calcio, ecc.*) **to s. twice** (*o* **to s. two goals**), segnare due gol; segnare due volte; fare una doppietta • **to s. a win**, ottenere una vittoria; vincere • (*calcio*) **to s. the winning goal** (*o* **a winner**) **in the last minute**, segnare il gol della vittoria (*o* risolvere la partita) in zona Cesarini • (*calcio*) **to s. a win with a header**, segnare di testa; incornare in gol (*fam.*) • (*hockey g.*) **to s. with an open cage**, segnare a porta vuota • (*calcio, ecc.*) **to s. with an open goal**, segnare a porta vuota (*o* a portiere battuto) • (*di una squadra*) **to be able to s.**, riuscire a segnare; arrivare al gol: **At long last we were able to s.**, finalmente siamo arrivati al gol • **not to be able to s.**, non riuscire a trovare la strada del gol • (*calcio*) **They've scored somehow**, c'è scappato il golletto.

We can observe that in Ragazzini (1998) alphabetical order only applies to headwords and compounds, while phrases are embedded under the main lemma entries. By contrast, headwords, compounds as well as phrases are listed separately in Bonanno (1988). Not only does Ragazzini (1998) provide the Italian equivalents of the English terms, but on several occasions it also includes further information (e.g. **to score a run**, (*baseball*) realizzare un 'run'; fare un punto con un 'home run'; segnare un punto facendo il giro completo del campo fino alla casa base) (Note 7) and explanations (e.g., **In cricket, a run is scored whenever the two batsmen exchange their wicket positions on hit or passed balls without either being out**) accompanied by their translation (nel cricket, si segna un 'run' ogni volta che i due battitori fanno la corsa per il cambio di wicket dopo che la palla è stata battuta ovvero passata, purché nessuno dei due si trovi fuori dalla zona (*di norma un punto va a entrambi*)).

The second part of the dictionary, the Italian-English section, resembles Bonanno's (1988) structure, because it is less detailed and simply works as a glossary. Some entries are more complete than others, but when phrases are incorporated in the entries they only consist of basic ones. Longer expressions are listed separately. Consider, for instance, the rather concise entry for 'segnare', which only appears with the equivalent English verbs related to various disciplines (Table 4). This lemma is then followed by 88 distinct entries with different degrees of complexity, e.g. **segnare agevolmente** (*calcio, ecc.*) to score freely; **segnare con un'angolazione perfetta** (*calcio, ecc.*) to score with a perfect angle/diagonal shot; **segnare con un secco tiro ricurvo piazzando il pallone sull'interno del primo palo** (*calcio*) to hook sharply inside the near post, etc.

Table 4. Entry for 'segnare' in Ragazzini (1998)

segnare (*un nome, ecc.*) to put down; to write down: (*tenere segnati i punti*) to mark; to keep (the score); to score; to tally; (*fare: punti*) to score; to score up; to take; (*baseball, basket, cricket, hockey, polo, tennis, ecc.*) (*colpi, punti*) to hit, to shoot; to play; (*calcio, ecc.*) to deliver; to notch (*anche* to notch up); to prod home; to put in; to realize; to score; to kick a goal; to put the ball home; to claim; to get a goal; to hit a goal; to make a hit; to run in; to score a hit; to strike a goal; to shoot; to kick; to kick the ball in; to kick in; (*trasformare*) to convert; (*insaccare*) to net; (*hockey g.*) to net, to net the puck; (*indicare*) to point.

Interestingly, the expression **segnare il gol della vittoria in zona Cesarini** is present in Ragazzini (1998) too, although in the English-Italian section it appears alongside **segnare all'ultimo minuto** as the equivalent of *to score (the winning goal/a winner) in the last minute*. Not even Ragazzini (1998) however makes reference to the colloquial use of the expression.

The weaknesses of this dictionary reside in the fact that it is not corpus-based, that it does not systematically provide the pronunciation of words and that no illustrations are included. Semantically related items are not clearly disambiguated either. If we consider the pair of near-synonyms discussed above, i.e. *cheer* and *applause*, what emerges in Ragazzini (1998) is that *cheer* as a noun has a richer semantic contour (i.e., it also has the sense of 'acclamation' and 'incitement' in addition to indicating the 'shouts' of people cheering) than its Latinate counterpart *applause* (Table 5). However, they appear to be completely interchangeable in the examples provided: not only can we say *to receive much applause*, but also *to receive many cheers*; similarly, the sentence *The cheers of the crowd on the terraces went sky-high* could be rephrased as *The applause of the crowd...* without significantly altering the meaning. In other words, the different meaning nuances and the limits of substitutability of the two words are not clearly presented.

Table 5. Entries for *applause* and *cheer* in Ragazzini (1998)

applause, *n.* (*senza pl.*) applause; applause: **to receive much a.**, ricevere molti applausi.

cheer, *n.* **1** applauso; incoraggiamento *(dei tifosi)* **2** *(pl.)* – **cheers***,* applausi, acclamazioni, incitamenti; (le) grida; (il) tifo; (i) cori: **The cheers of the crowd on**

> **the terraces went sky-high**, il tifo della folla sulle gradinate era alle stelle.

Although Ragazzini (1998) was published ten years later than Bonanno (1988), it does not show attention to representing genders equally in its entries either: certain words or expressions, such as *winner*, *conqueror* and *top finisher* for instance, are still only translated with their masculine forms, i.e. 'vincitore', 'conquistatore' and 'primo arrivato', respectively. It has become common practice in modern dictionaries instead to include both sexes (Note 8).

2.3 Cambridge Word Routes – Dizionario Tematico dell'Inglese Contemporaneo (1995)

Although this is not an ESP dictionary and is mainly aimed at Italian EFL (English as a Foreign Language) learners, it includes several thematic sections about sport and games. It is thus worth examining how it is internally organized and in what way it differs from the other two resources examined above. From a macrostructural perspective, it is similar to both Bonanno (1988) and Ragazzini (1998), because it also consists of a more detailed English-Italian first part and of a second section with just two lists of alphabetically ordered words in English and Italian.

The first word group related to sport is a general one that does not include vocabulary about any discipline in particular (it lists words such as **play**, **exercise**, **score**, **competition**, **pitch**, **field**, **ground**, and so forth), while the following ones cover ball sports, athletics, water sports, gymnasium sports, outdoor sports, target sports and equestrian sports. They are followed by some other general sections subsuming vocabulary related to the topic of 'success', 'failure', 'reward', 'strength', etc. The dictionary includes a separate section on games with rather basic vocabulary (e.g., **play**, **toy**, **quiz**, **cards** and so on).

Since this dictionary is meant to be used in the context of teaching/learning, many grammatical notes are included in the entries, especially when there is a difference in use between a certain English word/expression and its Italian counterpart. The entry for **walk** in the equestrian sports section (Table 6), for instance, shows that the verb in English may also be used transitively (Note 9), a possibility that is ruled out in Italian.

Table 6. Entry for *walk* in *Cambridge Word Routes* (1995)

> **walk** *v* **1** *vi* andare al passo **2** *vt* andare [ogg: solitamente un cane o un cavallo] condurre al passo.

Although the various sport sections include only essential vocabulary, they sometimes entail encyclopedic components. The entry for **rounders**, for instance, explains what sort of game it is and where it is played (Table 7). Nevertheless, the information provided is often not sufficient to understand what exactly is being indicated, especially when, as in this case, there is no corresponding term in Italian (Note 10) and no illustrations are provided.

Table 7. Entry for *rounders* in *Cambridge Word Routes* (1995)

> **rounders** *snn* [gioco inglese simile al baseball, giocato spec. a scuola da squadre femminili] rounders.

Interestingly, some entries contain information about style: **ping-pong** is defined as "rather informal, not used in sports competitions". However, the fact that this discipline is essentially the same as **table tennis**, which is also present in the section on racket sports, is not mentioned. This is because there are no section-internal references.

All the lemmas included in the various sections also appear in the two indices at the end of the volume, together with IPA (International Phonetic Alphabet) symbols, so that learners can be guided in accessing correct British English pronunciation. The didactic orientation of the dictionary is also reflected in the use of drawings illustrating some of the entries, albeit rather simply and not systematically.

3. Online glossaries of sports and games terms

The web abounds with glossaries of sports and games terms. Many of them are monolingual English resources that include terminology related to specific disciplines. They tend to appear in blogs, on the webpages of language schools or translation agencies and present different degrees of accuracy and comprehensiveness. It is often not

possible to verify the competence of those who created them or to know what criteria were followed for the selection of the terms. The main feature of most online glossaries resides in the fact that they are essentially word lists, sometimes even without proper definitions or usage examples (Note 11). In other cases, however, they can be more detailed and include explanations, images and video clips (Note 12). Interesting resources may also be found on the webpages of amateur or professional sports organizations. The National Collegiate Athletic Association in the USA, for instance, has online multimodal glossaries of the various sports played across the country, in Puerto Rico and Canada: the women's volleyball section, just to make an example, contains a dictionary and a list of common terms accompanied by realistic images, animations, illustration videos and clips. The National Hockey League also has a page dedicated to key hockey terms, even though it is quite static and not enriched with multimodal features (Note 13). Similarly, several other sports associations, organizations and clubs have developed more or less exhaustive databases with relevant terminology for a certain area.

Online bilingual English-Italian glossaries are few and typically include only discipline-specific terms. Most of them have an amateur character and were not created by language experts or lexicographers. An example is the glossary of dressage, compiled by someone with an interest in equestrianism (Note 14). Other resources can be found on the webpage of an Italian freelance translator (Note 15) which, among other things, contains links to bilingual as well as multilingual glossaries of terms pertaining to the various sports disciplines played during the Olympic games of 2012 and 2016. Another interesting platform is the Italian Language and Culture blog FluentU, which has a sports section with essential vocabulary 'for athletes and fans' (Note 16). Not only are the most popular sports listed, but there is also a section with the recurrent phrases and sentences used to speak about sport, accompanied by audio files. Since the latter provide the pronunciation of words and expressions only in Italian, it is evident that the target audience are prevalently native English speakers, although Italian users and learners of English may also benefit from them. A similar but richer website is that of Berlitz USA, which lists '149 fun sports terms in Italian to help you kick language goals' (Note 17), grouped according to the various types of disciplines, i.e. outdoor sports, water sports, indoor sports, etc., and accompanied by lists of sports-related vocabulary, such as verbs indicating sports actions (e.g., *to shoot*, *to dribble*, *to dunk*) and other generic words (e.g., *athlete*, *team*, *coach*). Because Berlitz is a language instruction company, with a focus to develop conversational skills and cultural awareness, this section of the website also includes example conversations in Italian about sports-related topics and whilst playing sports, in addition to information about the Italian sports habits and traditions.

Lastly, recent years have seen the emergence of video glossaries too, such as the one made available by Collins Dictionary (Note 18), which however present very basic vocabulary and are meant to be used in the context of initial language teaching/learning. Due to their simplicity, they cannot be treated as proper lexicographic resources.

4. Creating a bilingual English-Italian e-dictionary of sports and games

The overview of the existing bilingual English-Italian resources conducted above has evidenced that there is room for improvement in this area of ESP lexicography. The main shortcoming of sports and games lexicography today, at least for the English-Italian language pair, is that there exists no complete or updated dictionary in electronic form. The creation of such e-dictionary undoubtedly represents a challenge as users now expect to find an added value in dictionaries compared to what is already available on the web. Artificial intelligence can indeed provide very accurate translations of words, phrases and even whole texts, leading to the perception that dictionaries are now less necessary than in the past. Modern lexicographers thus need to think of ways to enrich dictionaries with features that justify resorting to them instead of performing ad-hoc online searches.

In this section, a model e-dictionary entry is presented taking as an example the word *padel*, indicating a new sports discipline that has become very popular in the past few years.

Let us start by considering what the initial part of the entry could be like. Following the recent trends of online dictionaries, it would make sense to include not just the phonetic transcription of the lemma, but also audio clips to listen to its pronunciation. This is particularly relevant when there is a difference across varieties of English: in *padel* the stress falls on the first syllable in British and Australian English (/ˈpadl/ or /ˈpæd(ə)l/), but also on the second syllable in American English (/pəˈdɛl/). Short clips can be found on YouTube presenting the pronunciation peculiarities of certain words (Note 19).

The main advantage that e-lexicography offers is the possibility to incorporate hyperlinks within dictionary entries that allow us to have access to a plethora of information (Hargraves, 2020), which for obvious reasons of space could not be included in paper dictionaries. In addition to audio and video files for pronunciation lookups (Figure 1), the lemma *padel* may be associated to the broader category of 'racket sports' by means of a gloss linked to a Wikipedia page, for instance, where more details about the discipline can be found. Hyperlinks would have to be active not just by clicking on them, thus redirecting users to external pages, but also by simply hovering over them with the mouse (Figure 2). This option should allow the display of images and pop-up windows partially showing the contents of external resources.
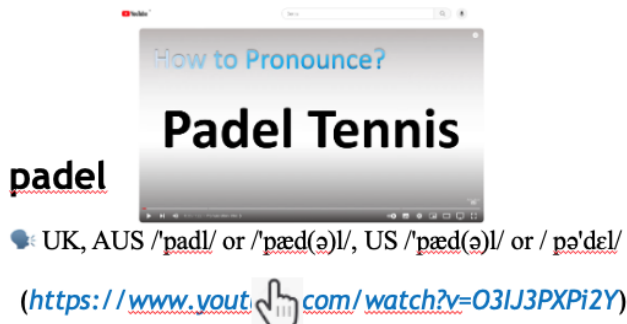
Figure 1. Pronunciation section for *padel*
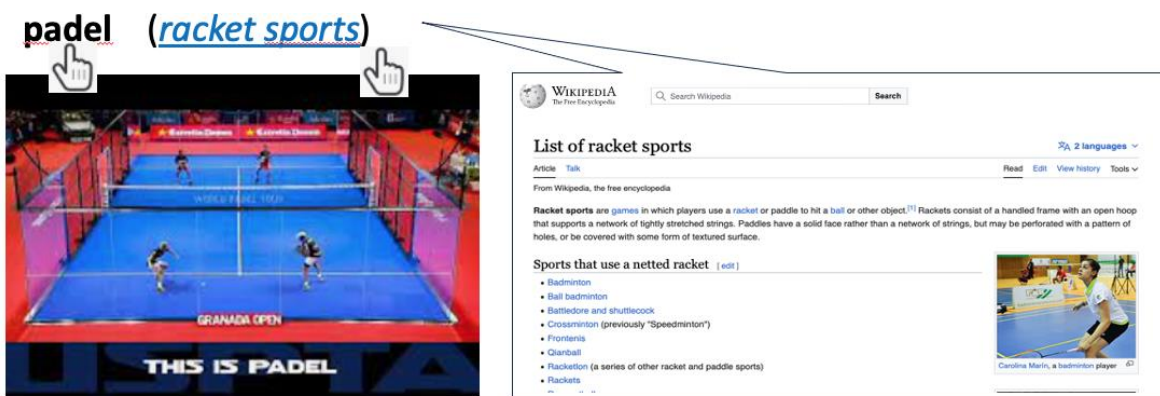
Figure 2. Image for *padel* and external link to English Wikipedia

The definition section of the entry should also come with the pronunciation of the word in the target language (and potentially with audio clips too) and with hyperlinks to external pages in Italian (Figure 3). This would eventually result in the creation of a bidirectional entry to be consulted to a certain extent by English- and Italian-speaking users alike.

Figure 3. Definition section for *padel* and external link to Italian Wikipedia

The phraseology section would have to include realistic examples from authentic sources and, again, hyperlinks to reliable external pages with additional details, e.g. about the rules of the game. A list of external links may be provided at the bottom of the page, in a manner similar to what happens on Wikipedia, so as to ultimately produce a dictionary-cum-encyclopedic e-entry with the most relevant linguistic and non-linguistic information about the term. Newspaper and magazine articles in both English and Italian may also be linked in a 'Find out more' section. Figure 4 shows what a complete entry for *padel* in a potential bilingual English-Italian e-dictionary of sports and games terminology may look like.

padel (*racket sports*)

🔊 UK, AUS /ˈpadl/ or /ˈpæd(ə)l/, US /ˈpæd(ə)l/ or / pəˈdɛl/ (*https://www.youtube.com/watch?v=O3IJ3PXPi2Y*)

padel 🔊 IT /ˈpadel/ (*sport della racchetta*), s.m., sport di derivazione tennistica praticato a coppie, con racchetta dal piatto rigido, su campo rettangolare all'aperto o chiuso da pareti su quattro lati.

- **padel court**, campo da padel; **padel rules**, **regole del padel**.

- **the ball hit the glass and went over the fence**, la palla ha colpito il vetro ed è uscita dal campo; **you have to hit/play the ball before the second bounce**, la palla deve essere colpita prima del secondo rimbalzo; **padel games are usually self-refereed**, di solito le partite di padel si giocano senza arbitro.

External links/Collegamenti esterni
International Padel Federation (FIP), https://www.padelfip.com
Federation of European Padel, https://www.padel-europe.org
Federazione Italiana Tennis e Padel, https://www.fitp.it

Find out more/Per saperne di più
The Guardian, https://www.theguardian.com/world/2022/feb/11/padel-italy-takes-off-no-sport-such-success-so-short-time
La Gazzetta dello Sport, https://www.gazzetta.it/fitness/allenamento/30-04-2021/pickleball-che-cos-come-si-gioca-guida-410603538673.shtml

Figure 4. Complete e-dictionary entry for *padel*

The creation of a multimodal and multimedial e-dictionary is a challenging task for a number of reasons. First, hyperlinks 'age' over time: the longer they exist on the web, the more they deteriorate and eventually become inactive or inaccessible. Therefore, they need to be constantly checked and substituted when necessary. This is a very time-consuming and laborious process that requires significant amounts of (financial) resources and efforts. Second, the availability of images and videos is usually subject to copyright, i.e. lexicographers are not free to utilize materials taken from other sources without obtaining prior permission. This explains why simple (black and white) drawings are sometimes used to illustrate lemmas. Figure 5 shows two images in The Free Dictionary (Note 20) for the word *abseil*.

Figure 5. Images for *abseil* in The Free Dictionary

If the use of uncopyrighted material represents a possible solution, it is also true that finding such material is not easy. Similarly, asking an illustrator to produce hundreds of pictures is not an easy option either and in any case the result would not be the same as using real photos. Third, new sports and games keep cropping up, a fact which also requires the constant updating and addition of entries. *Pop tennis*, for instance, which is not to be confused with *padel*, is not present in dictionaries; *footgolf* is only available in Wiktionary (Note 21); and so forth. Finally, do all sports and games have to be treated equally in terms of detail of the dictionary entry? It is obvious that the most popular disciplines, such as soccer, football and rugby, are likely to receive greater space; however, should synchronized swimming be treated less generously than boxing or kayaking? What about roller skating, chess, quoits, punting and weightlifting? This aspect opens up the question of how and to what extent different sporting cultures can be represented. If dictionaries should also have an edutainment function (Geeraerts, 2000), it thus makes sense to include as much information as possible in them. Because space is not a problem with e-dictionaries, it would be possible to add notes for certain culture-loaded terms, for instance. Consider the difference between the Italian bowling game of *bocce* and simple *bowling*. The former started as an inexpensive and popular game among the working class and it is usually associated with older adults or retired people, typically men, meeting at local 'circoli' (i.e., anti-Fascist gathering points); the latter instead usually refers to pin bowling and is practiced everywhere and by everyone. Another example is horse racing in Italy, where it also comprises certain traditional games, such as the Palio race in Siena, which is not just a sports competition, but rather a year-long

event that is very much part of the local Tuscan culture. These are important facts that need to be addressed for a better understanding of what is behind a term.

5. Conclusions

The bilingual English-Italian dictionaries of sports and games terminology currently available on the market are only in paper form; they have not been recently updated and are thus rather out of date. More dynamicity and an electronic format in line with modern lexicography are called for. The present paper has tentatively presented a possible improved e-dictionary entry showing the kind of features that need to be present in order to enrich users' browsing experience and make it effective. The difficulties of creating a multimodal and multimedial entry have also been discussed. They mainly concern the availability of authentic materials and having to tackle copyright issues.

In an era where the Internet offers an astonishing abundance of resources, dictionaries necessarily need to incorporate rapidly accessible information all in one page that would otherwise have to be retrieved with multiple searches. This should be the main strength of e-lexicography today. The future holds both challenges and promises in this field. While on the one hand artificial intelligence is often used as an alternative to dictionaries, it could also help lexicographers to create dictionary entries more rapidly. AI-generated entries can indeed be produced using ChatGPT. The path to replacing human labor with automatic process, however, is still unclear. It seems plausible to say that some form of human intervention and post-editing will continue to be necessary in the coming years, at least to avoid plagiarism and copyright infringement.

References

Binder, T. & Brasse, M. (1998). *Football dictionary. Dictionnaire du football*. Zürich: FIFA.

Bonanno, V. (1988). *Dizionario dello Sport e di Medicina Sportiva inglese-italiano, italiano-inglese.* Roma: Edizioni Mediterranee.

Cambridge Word Routes – Dizionario Tematico dell'Inglese Contemporaneo (1995). Cambridge: Cambridge University Press.

Fuertes-Olivera, P.A. (2018). Introduction: lexicography in the Internet era. In Fuertes-Olivera, P.A. (Ed.), *The Routledge Handbook of Lexicography* (pp. 1–15). Abingdon, Oxon: Routledge.

Furiassi, C. & Gottlieb, H. (2015). *Pseudo-English. Studies on False Anglicisms in Europe*. Boston & Berlin: Mouton de Gruyter.

Geeraerts, D. (2000). Adding Electronic Value. The electronic version of the Grote Van Dale. In: Heid, U., Evert, S., Lehmann, E. & Rohrer, C. (Eds.), *Proceedings of the Ninth EURALEX International Congress, Stuttgart, Germany, August 8th - 12th, 2000* (pp. 75–84). Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung.

Hargraves, O. (2020). Electronic Dictionaries. In S. Ogilvie (Ed.), *The Cambridge Companion to English Dictionaries* (pp. 207-218). Cambridge: Cambridge University Press.

Jackson, H. (2018). English Lexicography in the Internet era. In Fuertes-Olivera, P.A. (Ed.), *The Routledge Handbook of Lexicography* (pp. 540–553). Abingdon, Oxon: Routledge.

Pons, 2006. *PONS Überlebenswortschatz Fußball. Deutsch, Englisch, Französisch, Italienisch, Niederländisch, Portugiesisch, Spanisch*. Stuttgart et al.: Klett.

Pulcini, V. (2008). Anglicisms in the 2006 Winter Olympic Games. In R. Fischer & A. *Pułaczewska (Eds.) Anglicisms in Europe: Linguistic Diversity in a Global Context* (pp. 140–158). Newcastle-Upon-Tyne: Cambridge Scholars Publishing.

Ragazzini, G. (1998). *Dizionario dello sport. Inglese-italiano, italiano-inglese*. Bologna: Zanichelli.

Sirges, H. (1980). *Elsevier's football dictionary: English/German, German/English*. Amsterdam &New York: Elsevier.

Taborek, J. (2020). Multilingual LSP dictionary. Lexicographic conception of a dictionary of football language. In M. J. Domínguez Vázquez et al. (Eds.), *Studies on Multilingual Lexicography* (pp. 213–228). Boston & Berlin: Mouton de Gruyter.

Wehlen, R. (1976). *Regeln und Sprache des Sports: Fussball, Handball, Korbball, Hockey, Tennis, Golf, Polo, Basketball, Wasserball und alle anderen Ballspiele*. Mannheim et al.: Bibliographisches Institut/Dudenverlag.

Yıldırım, K. (2006). *Kauderwelsch Dictionary. Fußball – Football – Calcio – Fútbol – Futebol – Futbol. In 7 Sprachen: Deutsch, Englisch, Français, Italiano, Español, Português, Türkçe.* Bielefeld: Reise-Know-How Verlag Peter Rump.

Notes

Note 1. The modern game of tennis originated in England in the 19th century as lawn tennis; rugby was born in the city of Rugby, Warwickshire in the first decades of the same century; rowing started on the river Thames in London in the same period, etc.

Note 2. While in some cases anglicisms remain unchanged in Italian (e.g., *cricket*, *polo*, *tennis*), there are also several instances of modifications (e.g., *snowboarding* may be referred to as 'snow' as in "Non scio ma faccio snow", i.e. I don't ski, but I snowboard), which sometimes also produce false anglicisms (e.g., the use of 'basket' to indicate *basketball*). See Furiassi & Gottlieb (2015) for an in-depth examination of pseudo-English.

Note 3. This is my translation of a paragraph in the 'Presentation' section.

Note 4. Although bilingual dictionaries are prevalently repositories of words with their translations, the addition of some encyclopaedic information is useful to facilitate comprehension and to enrich what are otherwise rather sterile entries.

Note 5. This may be the case of medical terminology, e.g. It. *frattura composta*, En. *compound fracture*, whose definitions would be unnecessary for sports doctors.

Note 6. Incidentally, this dictionary lists all the items alphabetically, with no nesting, thus avoiding textual condensation and facilitating accessibility. At the same time, however, such procedure deviates from the standard way of organizing lemma entries in dictionaries and it is perceived as a rather amateur lexicographic technique.

Note 7. "Fare un punto con un 'home run'" or "segnare un punto facendo il giro completo del campo fino alla casa base" may be translated as 'to score a point by going all the way around the field to home plate'.

Note 8. The entry for *winner* in the *Collins Online English-Italian Dictionary* (https://www.collinsdictionary.com/it/dizionario/inglese-italiano/winner), for instance, includes the gloss *gen* to indicate that in Italian it is a gendered word and thus provides both the masculine and feminine form, i.e. "vincitore/trice". Last access: 20 December 2023.

Note 9. The part in square brackets in the entry indicates that the object of walking is usually a dog or a horse.

Note 10. The part in square brackets in the entry indicates that it is "an English game similar to baseball, played especially at school by female teams".

Note 11. EnchantedLearning.com (last access: 20 December 2023), for instance, consists of 317 items grouped alphabetically without definitions or examples.

Note 12. The website of ATS Language, a Hungarian translation agency, includes a useful page (https://www.ats-group.net/glossaries/glossary-lexicon-sports.html) with many links to external glossaries and lexicons of several sports and games (last access: 20 December 2023).

Note 13. https://www.nhl.com/info/hockey-glossary (last access: 20 December 2023).

Note 14. https://web.archive.org/web/20090301234627/http://www.geocities.com/sturmele/Tesi%20inglese.htm (last access: 20 December 2023).

Note 15. https://www.raffaellalippolis.com/sport-e-olimpiadi-glossari-e-risorse-terminologiche (last access: 10 December 2023).

Note 16. https://www.fluentu.com/blog/italian/italian-sports-vocabulary (last access: 10 December 2023).

Note 17. https://www.berlitz.com/blog/sports-italian (last access: 10 December 2023).

Note 18. https://www.youtube.com/watch?v=VyMBpJf4zSg (last access: 20 December 2023).

Note 19. This is also the case for *padel*: https://www.youtube.com/watch?v=O3IJ3PXPi2Y (last access: 20 December 2023).

Note 20. https://www.thefreedictionary.com/abseil (last access: 20 December 2023).

Note 21. https://en.wiktionary.org/wiki/footgolf (last access: 20 December 2023).

Copyrights

Exploring the lexis of art through a specialized corpus:
A bilingual Italian-English perspective

Antonella Luporini[1]

[1] Department of Modern Languages, Literatures and Cultures, University of Bologna, Bologna, Italy

Correspondence: Antonella Luporini, University of Bologna, Department of Modern Languages, Literatures and Cultures, via Cartoleria, 5, 40124 Bologna, Italy. Tel: +39-051-20 97 240 E-mail: antonella.luporini@unibo.it

Abstract

This study presents an application of a specialized corpus, including texts specifically related to art and cultural heritage, to the analysis of artistic vocabulary in a bilingual (Italian-English) perspective, focusing on the Italian lemmas *opera*, *figura* and *disegno* and their English translation equivalents. The starting point is the Italian corpus that is being developed under the research project *Lessico multilingue dei beni culturali* ('Multilingual art and cultural heritage vocabulary', LBC), available online in open access through NoSketchEngine. First, a lemmatized nounlist ordered by frequency of occurrence is extracted from the corpus, leading to the selection of the above-mentioned focus words, in view of both their frequency and status as technical terms within the domain of art (though exhibiting different levels of technicality). These are further investigated by extracting collocates and KWIC concordances, leading to the identification of several specialized collocations and domain-specific senses. The analysis subsequently moves from corpus to dictionary, exploring the extent to which the patterns emerging from corpus investigation are accounted for in the entries for *opera*, *figura* and *disegno* in four Italian-English bilingual dictionaries. From this viewpoint, the study also aims to show how specialized corpus data can be used for the extraction of collocations, terms, and context-specific word senses, which may in turn be used both to enrich the information provided by currently available general dictionaries, and to work towards the creation of a large-scale specialized bilingual dictionary, which is non-existent to date.

Keywords: art lexis, bilingual lexicography, collocations, polysemy, specialized corpus, Italian-English translation

1. Introduction

The specialized language of art is going through a process of visibilisation, thanks to the growing number of public initiatives and online resources fostering the preservation of cultural heritage and promoting sustainable tourism, such as the UNESCO World Heritage List, the European Union's Heritage Days, or the events organized by the National Trusts for England, Wales and Northern Ireland, Scotland and Italy, to name a few. Domain-specific vocabulary is consequently being used in an increasing variety of genres and contexts – promotional/informative web contents, webinars and museum virtual tours, alongside more traditional resources such as tourist guidebooks and brochures – all, by their very nature, addressing a multilingual audience.

At the same time, specialized bilingual lexicographic resources that could support, e.g., translators, web content creators and visitor bureaux in the multilingual description of art and cultural heritage are still scarce (Flinz, 2023). With reference to the Italian-English language pair, while several monolingual dictionaries, glossaries and thesauri of art terms are available (Note 1), a bilingual resource with comparable levels of comprehensiveness and specialization, to date, does not exist (Note 2).

Specialized corpora of texts from the domain of art and cultural heritage would also represent a key resource for translation and lexicographic purposes. Indeed, the fact that corpus data can reliably support different decision-making processes involved in the creation and updating of dictionaries is a given of modern lexicography: among these, defining lemma-lists, choosing relevant examples and retrieving complex information (lexical-semantic, pragmatic, collocational/colligational, connotational etc.) to be included in dictionary entries (see, e.g., Berti & Pinnavaia, 2012; Faaß, 2018; Rundell, 2018). However, until a few years ago, there were no freely accessible corpora of this kind (Flinz 2023).

With a view to filling this gap, back in 2016 the research unit *Lessico multilingue dei beni culturali* ('Multilingual art and cultural heritage vocabulary', henceforth LBC) at the University of Florence (Italy) launched a project for the creation of a large databank of texts related to the domain of art and cultural heritage, belonging to different text-types (technical, literary, informative), and in different languages – Italian, English, French, German, Russian and Spanish (see Farina & Nicolás Martínez, 2020). The project was subsequently extended to other universities, among which Bologna, Milano Statale, Paris 8 and Pisa. At its current stage of development, it has led to the

creation of six specialized comparable corpora (one for each of the project languages), including at least 1 million words, which are freely accessible online through NoSketch Engine, a free version of the corpus management system Sketch Engine (Note 3). This resource remains one of a kind to date, both in terms of coverage and accessibility. The ultimate goal of the project, currently work-in-progress, is the creation and open-access publication of a corpus-based, specialized multilingual dictionary, made up of single monolingual dictionaries with dynamic connections among the entries in the various languages, and of several parallel corpora, where source texts in the project languages will be aligned with their available translations (see Zotti, 2017; Farina & Billero, 2020; Flinz, 2023).

Against the above background, we will now move on to describe an application of the currently available LBC corpora to the analysis of the lexis of art in an Italian-as-source/English-as-target perspective. This contribution qualifies as a corpus-driven study (Tognini-Bonelli, 2001), in that the choice of items to be investigated and the analysis/identification of patterns are fundamentally driven by corpus criteria, as explained in the 'Method' section below. Focus is on the Italian lemmas *opera*, *figura* and *disegno*, and related translation equivalents in English. By comparing corpus evidence with the information provided by several bilingual dictionary entries, this study ultimately aims to offer new perspectives on the lexicographical choices made with reference to these words, from the specific perspective of the translation of art vocabulary.

2. Method

Since it was decided, at the study design stage, to focus on Italian as the source language, the Italian LBC corpus (Note 4) provides the starting point for this study. This corpus currently consists of approximately 1,186,000 word tokens, mostly from two of the text-type categories envisaged by the project: technical and literary (64% and 28.3%, respectively). A key source of texts for this corpus is Giorgio Vasari's *Le vite*: a collection of biographies of the most eminent Italian painters, sculptors and architects, written in the 16th century. *Le Vite* is a foundational text for the description of Italian Renaissance art and for art history more generally (Note 5); due to its comprehensiveness and complexity, it may be treated as a small specialized corpus in its own right (Luporini, 2023).

The analysis took place through the steps listed below.

2.1 Nounlist

The first step of research involved retrieving a wordlist from the Italian LBC corpus, with the aim of identifying lexical candidates to be examined. A wordlist is a list of all the words included in the corpus, ordered by raw frequency of occurrence; wordlists can also be lemmatized and annotated with part-of-speech information (see Baker, Hardie & McEnery, 2006, p. 169).

Nouns are particularly important in the creation of technical vocabularies, as indicated by the etymology of the word 'nomenclature' itself: "(from the Latin *nomen calare*) first appeared in French and English at the beginning of the 16th century, with the meaning of 'glossary' or '*list of names*'" (Rey, 1995, p. 11; emphasis added). Taking this into consideration, a wordlist tailored to the purposes of this study was generated, including only the nouns that appear in the corpus, in a lemmatized form: we shall refer to this as a 'nounlist'. This was done by choosing 'find tags matching regular expression (regex)' in the advanced wordlist options, filling in the regex slot with the tag NOUN.* (Note 6), and, finally, selecting 'display results as lemmas'. This procedure, which was devised as a 'shortcut' to the specialized words included in the corpus, also has the advantage of automatically eliminating from the list all the grammatical words that, due to their frequency, typically rank very high independently of the text-type, but only create 'noise' when analysis is oriented towards content words.

From the final output, to be illustrated in Section 3, the three items *opera*, *figura* and *disegno* were selected as foci for this study, based on their frequency and status as specialized vocabulary items.

2.2 Collocations and KWIC concordances

The second step of the analysis involved examining the behavior of the focus words emerging from the previous step through collocations and KWIC concordances. Collocates are words that tend to co-occur with a node word in a significant way in a corpus (see Baker, Hardie & McEnery, 2006, pp. 36 ff.). NoSketch Engine – just like its fully functional counterpart, Sketch Engine – provides various options for collocate identification. For this study, the collocational window (the span within which the system will look for collocates) was set to -3/+3 – from three words to the left to three words to the right of the node word; this is narrower than the 'traditional' -5/+5 window, but is supposed to yield more precise results in terms of strength of association (cf. Bartsch & Evert, 2014, p. 57). Candidate collocates were retrieved in lemma form and ordered by LogDice score, a statistical measure of the strength of collocational association (Note 7). One problem with this step was that the lists of collocates thus generated were skewed by an overabundance of grammatical words, and even punctuation marks, which occupied the top positions without being particularly revealing (Note 8). For this reason, it was decided to work on a significant portion of the collocate lists, taking into account all the candidates with a LogDice value higher than

7.0. Collocations worth further examination within the context and objectives of this study were then singled out. Concretely, this was done by 'moving back and forth' between the collocate lists and the related KWIC concordances, showing the collocates in their original, extended (sentence-length) co-text (see Baker, Hardie & McEnery, 2006, pp. 42 ff.).

2.3 Looking at dictionary entries through the lens of the corpus

The final step of the analysis involved moving from corpus to dictionary, with a view to assessing if, and to what extent, the specific collocations and word senses resulting from 2.2 above are included in the entries for our focus words. To this end, a set of four bilingual Italian-English resources was identified. This included two renowned free online dictionaries, the Collins Italian-English Dictionary online and the Cambridge Italian-English Dictionary online (Note 9), and the digital editions of two authoritative dictionaries, both requiring subscription: the Oxford Italian-English Dictionary and *Il Ragazzini* Zanichelli – undoubtedly a best-seller in Italy, featuring "the richest and most up-to-date range of aids for Italian students of English" (Iamartino, 2019, p. 149) (Note 10). This set is necessarily limited, due to the qualitative nature of this study, but it still aims at being representative of different publishing/lexicographic traditions (UK- and Italy-based companies) and also of different user needs. In fact – in the absence of a specialized bilingual dictionary focusing exclusively on the language of art, and with the growing demand for multilingual texts in this sector (cf. the Introduction) – it was hypothesized that free online dictionaries like Collins and Cambridge are more likely to be consulted by non-professional translators than Oxford or Zanichelli.

3. Results and discussion

3.1 Nounlist

The first step that will be illustrated is the choice of the focus words for this study. Taking into account the upper part of the LBC Italian corpus nounlist, showing the most frequent noun lemmas in the corpus (Table 1), the three interconnected items *opera*, *figura* and *disegno* were selected, being interpreted as items belonging to art vocabulary.

Table 1. LBC Italian corpus wordlist, top ten noun lemmas by frequency of occurrence

#	Noun (lemma)	Frequency	#	Noun (lemma)	Frequency
1	cosa	4,925	6	anno	2,063
2	**opera**	4,081	7	modo	1,939
3	tempo	2,356	8	parte	1,762
4	mano	2,324	9	**disegno**	1,647
5	**figura**	2,100	10	città	1,635

With reference to their being interconnected, *opera* may be taken as the superordinate, with *figura* and *disegno* as co-hyponyms (i.e., *figura* and *disegno* as 'types of' opera). However, between *opera* and *disegno*, on the one hand, and *figura*, on the other, a lexical relation of meronymy could also be found, depending on the context (i.e., *figura* as 'part of' an *opera* or *disegno*). At the same time, it must be noted that our focus words also instantiate different levels of *technicality*, understood as "the degree to which a term is specialized and exclusively used by experts in a domain" (Hätty *et al.*, 2020, p. 2883; see also Dima, 2012, pp. 95-96). *Disegno* – even though it is commonly used in general language, and its meaning has been metaphorically extended to different domains – still enjoys full status as an element of artistic terminology, as also highlighted by its etymology (Note 11). In fact, *disegno* is also the only lexical unit from this set having a dedicated entry in an Italian-language art encyclopedia that was consulted as a reference (*La nuova enciclopedia dell'arte Garzanti*, 1986). *Figura* is also etymologically linked to the domain of arts and crafts, broadly speaking (Note 12), but the connection is arguably more opaque in this case, probably also as a consequence of a broader process of polysemization. Similar considerations can be made with reference to *opera*, the item with the lowest level of technicality in this set (Note 13). By the same token, *mano* and *modo*, which had also been originally identified as potentially relevant items, were eventually put aside for a future study, after looking at their occurrences in KWIC concordances, and noticing that the more technical senses 'layer of color' and 'artistic manner' are not predominant in the corpus. All in all, the fact that technical or semi-technical terms rank so high may also be taken as a positive sign concerning the quality of this specialized corpus.

3.2 Collocations and KWIC concordances

3.2.1 *Opera*

Within the collocate list for *opera* generated with the criteria described in Section 2.2 above (Note 14), the lemmas with a LogDice value higher than 7.0 correspond roughly to the first 120. From this set, two candidate collocates worth further investigation were singled out: *bocca* (no. 62) and *fresco* (no. 103). The related concordance lines show that *opera*, on the one hand, and *bocca* and *fresco*, on the other, occur together within the expressions *bocca d'opera* (35 hits) and *opera in/a fresco* (14 hits).

Bocca d'opera is a lexicalized expression, commonly found as a single word in present-day Italian, which refers to the part of the stage in a theatre that is framed by the proscenium arch. Authoritative Italian dictionaries such as *Zingarelli*, *Sabatini-Coletti* and *Devoto-Oli* explain it by establishing a correspondence with *boccascena* ('proscenium'), even though, according to the *Manuale della lingua teatrale* by Papiol (1909), the two terms actually refer to different parts of the front stage, *boccadopera* being a specific part of the *boccascena* (Note 15). The multiword expression *bocca d'opera* in the LBC corpus is found within a single text: Giordani's *Intorno al Gran Teatro del Comune e ad altri minori in Bologna* (1855).

As for *opera in fresco* and *opera a fresco* (lit. 'work in fresco'), corpus data point to them as virtually exclusive to *Le Vite*, where Vasari uses them in place of *affresco* ('fresco'). In fact, *affresco* as a single word in Italian appears for the first time in the 19th century, from the bigram (*dipingere*) *a fresco*, of which *in fresco* is a less frequent variant. Indeed, a search for *affresco* and *affreschi* in the Italian LBC corpus yields only results from later texts, such as Gualandi's *Tre giorni in Bologna o Guida per la città e i suoi contorni* (1865), or Baraldi's *Alla scoperta dei segreti perduti di Bologna* (2016).

3.2.2 Figura

As was the case with *opera*, in the collocate list for *figura* (Note 16), the candidates with a LogDice value higher than 7.0 amounted to roughly 120. Five were shortlisted in this case: *storia* (no. 3); *tondo* (no. 12); *mezzo* (no. 18); *rilievo* (no. 23), *fresco* (no. 27) and *terra* (no. 69). By looking at the related concordance lines, several specialized expressions were identified, all occurring in the corpus as part of Vasari's *Le Vite*: these are listed below, together with explanations and English translation equivalents.

Storia di/in figure (31 hits), referring to the painted or sculpted representation of an event made up of various parts. En. 'scene with figures'.

Figura tonda (35 hits), denoting well-rounded sculpted or carved figures, in full relief. As Vasari himself explains in the *Introduction to the three arts of design* that opens *Le Vite*, "[s]uch figures [those showing due proportion, grace, design and perfection] we call figures 'in the round', provided that all the parts appear finished, just as one sees them in a man, when walking round him" (Maclehose & Brown, 1907, p. 147). En. 'figure (carved/cast) in the round', 'round figure', or 'figure in full relief'.

Mezza figura (25 hits), referring to the representation (painting, drawing, or sculpture) of a man or woman from head to waist. En. 'half-length figure'.

Figura di (*basso/mezzo/gran/tondo*) *rilievo* (23 hits), denoting sculpted figures, landscapes, or architectural/ornamental elements that emerge from the background in a more or less neat way. En. 'figures in (low/half/strong/full) relief'.

Figura in/a fresco (26 hits), referring to 'frescoed' figures. En. 'figure in fresco'.

Figura di terra (15 hits), i.e., sculpted figures made with *terra* in the sense of clay. En. 'figure in clay'.

3.2.3 Disegno

The candidates with LogDice higher than 7.0 in this case amounted to roughly 110 (Note 17). From this subset, the following were selected for further investigation: *modello* (no. 4), *invenzione* (no. 5), *grazia* (no. 8), *cartone* (no. 11), *ordine* (no. 15) and *giudizio* (no. 31), all occurring within nominal group complexes of the kind '*disegno* and x': *disegno e modello* (or *modello e disegno*, 41 hits); *disegno e invenzione* (or *invenzione e disegno*, 33 hits); *disegno e grazia* (or *grazia e disegno*, 28 hits); *disegno e cartone* (or *cartone e disegno*, 21 hits); *disegno e ordine* (or *ordine e disegno*, 27 hits) and, finally, *disegno e giudizio* (or *giudizio e disegno*, 13 hits). The related concordance lines were, once again, almost exclusively extracted from *Le Vite*.

As can be noted, these expressions cannot be said to realize specialized, terminologically relevant collocations (as was the case with *opera* and *figura*); their relevance for the present study rather lies in the fact that they emphasize the inherent polysemy of the word *disegno* in Vasari's masterpiece, already shown by previous contrastive translation studies on other language pairs (Carpi & Pano Alamán, 2019, on Italian and Spanish; Ballestracci, 2023, on Italian and German).

In fact, the analysis of the above-mentioned expressions in context by means of concordance lines points to three main senses of the word *disegno*, which emerge neatly from its association (in some cases, contrast) with the collocates being considered, and which can be put on a cline, from the more concrete to the more abstract: (i)

disegno as 'graphic representation' resulting from the act of drawing; (ii) *disegno* in the sense of 'preparatory draft', but also 'project', 'plan'; (iii) by extension, *disegno* as 'skill', 'talent', also 'inventive ability'. This breakdown obviously also impacts on the choice of an appropriate translation equivalent. Through the following concordance lines, we illustrate the difference between senses (i) and (ii) in relation to the collocation *invenzione e disegno*, by also comparing the Italian source text with De Vere's *Lives of the Most Eminent Painters, Sculptors and Architects* (1912-1915), which remains to date the only extant complete English translation of *Le Vite*. In (1), *disegno* basically refers to the (positively evaluated) final result, and is translated as 'drawing' by De Vere. In (2) and (3), *disegno* stands for 'skill', 'ability' and is translated by De Vere as 'draughtsmanship' and 'design', respectively.

(1) Sonovi oltre ciò molte grottesche e altre cose lavorate di chiaroscuro simili al marmo e fatte stranamente con **invenzione e disegno** bellissimo. (*Vita di Filippo Lippi pittore fiorentino*)

Besides this, it contains many grotesques and other things wrought in chiaroscuro to resemble marble, executed in strange fashion with **invention and** most beautiful **drawing**. (De Vere, 1913, Vol. IV, p. 8)

(2) Aveva Cosimo, sentendosi debole **d'invenzione e di disegno**, cercato di occultare il suo deffetto (*Vita di Cosimo Rosselli*)

Cosimo, feeling himself weak in **invention and draughtsmanship**, had sought to conceal his shortcomings (De Vere, 1912, Vol. III, p. 189)

(3) ella sarebbe certo bellissima tra tutte le cose d'Andrea: al quale, se la natura avesse dato gentilezza nel colorire come ella gli diede **invenzione e disegno**, egli sarebbe veramente stato tenuto maraviglioso. (*Vita di Andrea Dal Castagno e Domenico Veneziano*)

it would certainly be the most beautiful of all the works of Andrea. And if Nature had given grace of colouring to this craftsman, even as she gave him **invention and design**, he would have been held truly marvellous. (De Vere, 1912, Vol. III, p. 100)

At the same time, *disegno* as 'draft', 'project' or 'plan' – to be technically distinguished from a 'model' or a 'cartoon' – may also be rendered through 'design' in English, as shown by the following concordance lines related to *disegno e modello* and *disegni e cartoni*:

(4) egli, secondo ch'io truovo, fece il **disegno e modello** del palazzo de' Governatori della città d'Ancona (*Vita di Margaritone*)

he, according to what I find, made the **design and model** of the Palazzo de' Governatori in the city of Ancona (De Vere, 1912, Vol. I, p. 66)

(5) oltre a molti **disegni e modelli** che vi fece di abitazioni private e publiche (*Vita di Michelozzo Michelozzi*)

besides many **designs and models** that he made for private dwellings and public buildings (De Vere, 1912, Vol. II, p. 261)

(6) e con quanta diligenza esso Puntormo conducesse a perfezzione e ponesse ottimamente in pittura i **disegni e 'cartoni** di Michelagnolo (*Vita di Jacopo da Pontormo*)

with what diligence Pontormo carried to completion and executed excellently well the **designs and cartoons** of Michelagnolo (De Vere, 1914, Vol. VII, p. 172)

All these senses of *disegno* are likely to be found also in contemporary texts describing works of art or promoting cultural heritage, and are, therefore, relevant also from the perspective of bilingual lexicography and translation. Other 'nuances' of meaning of *disegno* – and potential translation equivalents – might be unveiled by future studies comparing *Le Vite* with different English translations through a wider and more systematic approach. This will be made possible by the implementation of parallel corpora, which is on the LBC project agenda (see Section 1).

3.3 From corpus to dictionary

The last stage of the analysis involved checking the entries for *opera*, *figura* and *disegno* in the Italian-English section of the bilingual dictionaries listed in Section 2.3, in order to look for the specialized collocations and multiple senses of the focus words that emerge from the corpus. Table 2 summarizes the main findings.

Table 2. Bilingual dictionary entries for *opera*, *figura*, *disegno*: presence (✓) or absence (✗) of corpus findings

Search item	Collins Ita-En Dictionary online	Cambridge Ita-En Dictionary online	Oxford Ita-En Dictionary	*Il Ragazzini* Zanichelli
Bocca d'opera	✗ [Only *boccascena* as separate entry > 'proscenium']	✗	✗ [Only *boccascena* as separate entry > 'proscenium']	✓ [*Boccadopera* present as separate entry > 'proscenium']
Opera in/a fresco	✗	✗	✗	✗
Storia di/in figure	✗	✗	✗	✗
Figura tonda	✗	✓ > 'round figure' [*figura rotonda*, not indicated as 'Art']	✓ > 'round figure' [*figura tonda*, not indicated as 'Art']	✗
Mezza figura	✓ > 'half-length portrait'	✗	✗	✓ > 'half-length portrait'
Figura di (basso/mezzo/gran/tondo) rilievo	✗	✗	✗	✗
Figura in/a fresco	✗			
Figura di terra	✗	✗	✗	✗
Disegno as 'graphic representation' – sense (i)	✓ > 'drawing'; (*schizzo*) 'sketch'	✓ > 'drawing'; 'sketch'; 'design'	✓ > 'drawing'; 'picture'	✓ > 'drawing'; (*schizzo*) 'sketch'
Disegno as 'draft', 'project', 'plan' – sense (ii)	✓ > 'outline'; 'plan'; 'project' [all indicated as 'figurative']	✓ > 'sketch'; 'outline'; 'design'	✓ > 'design'; 'scheme'	✓ > 'outline'; 'sketch'; 'design'; 'plan'
Disegno as 'skill', 'talent', 'inventive ability' – sense (iii)	✗	✗	✗	✓ > 'draftsmanship' 'draughtsmanship';

As can be seen from Table 2, none of the consulted dictionaries includes all the information emerging from our corpus investigation. This is perhaps not surprising, considering that these are general dictionaries. However, given the paucity of specialized reference resources currently available for the translation of art lexis – and the consequent need for expert and non-expert translators to rely on general resources – a few additions based on corpus evidence would be desirable. These concern especially the entries for *figura* and *disegno*. The findings related to *bocca d'opera* and *opera in fresco*, which are archaic forms only weakly linked to *opera* in present-day language, are probably not as relevant for general dictionaries as they could be for the creation of a specialized comprehensive dictionary.

Regarding *figura*, it would be worth adding the specialized collocations/terms *storia di/in figure* ('scene with figures') and *figura di terra* ('figure in clay') that are absent from all the examined entries, especially considering that the words *storia* and *terra* may be misleading for Italian speakers who are not versed in art terminology and may interpret and translate them as 'story' and 'earth'/'soil', respectively. *Mezza figura* could also be incorporated

in the Cambridge and Oxford dictionaries, while the superordinate 'half-length figure' could be included along with 'half-length portrait' in Collins and *Il Ragazzini*, thereby more explicitly including carved realizations, in line with corpus evidence. As for *figura tonda*, the translation equivalent 'figure in the round' could also be added and explicitly signalled as belonging to the field of art. In the case of *figura di rilievo*, since the bigram *in rilievo*, translated as 'in relief', can be found under *rilievo* in all the dictionaries under scrutiny, it would probably be unnecessary to also mention it in the entry for *figura* (although it must be noted that Cambridge only mentions 'in relief' with specific reference to the globe or a map). However, the entries for *rilievo* could also accommodate the collocations *di basso/mezzo/gran/tondo rilievo*, with the respective translation equivalents 'in low'/'half'/'strong'/'full' relief. Currently, only Collins and Ragazzini have both 'high-relief' and 'bas-relief'; Cambridge only mentions 'bas-relief' and Oxford does not give any specification. As for *figura in/a fresco*, the observations made above with reference to *opera in/a fresco* remain valid.

Concerning *disegno*, while all the dictionaries account for senses (i) and (ii) as emerging from the corpus analysis presented above, sense (iii) can only be found under this lemma in *Il Ragazzini*, translated as 'draughtsmanship' (AmE 'draftsmanship'). Cambridge, under the label "*arte di disegnare*", provides translations that appear to apply more to the subject taught in schools than to a type of art ('drawing', 'design', 'graphic arts') – therefore, these were not included in Table 2. A more comprehensive account of the complex polysemy of this word would be provided by the other dictionaries if they acknowledged this sense as well. Additional corpus-assisted studies are needed to provide evidence that may shed light on the contextual difference between apparently interchangeable equivalents like 'design' and 'sketch', on the one hand, and 'design' and 'cartoon', on the other. For instance, Cambridge, differently from the other dictionaries, includes 'design' among the possible equivalents for our sense (i); Cambridge and *Il Ragazzini* mention 'sketch' among those for sense (ii); *Il Ragazzini* translates *disegno preparatorio* as 'preparatory drawing' or 'cartoon', but this does not seem to fit in well with corpus evidence showing the existence of the collocation *disegni e cartoni* in Vasari (see Section 3.2.3). These remain open questions to be investigated with the indispensable aid of parallel corpora that comprise more recent or contemporary source texts and translations.

4. Conclusions and outlook

This study has presented a possible application of the specialized corpora that are currently being developed as part of the LBC project to the analysis of the lexis of art for bilingual lexicographic purposes. More specifically, it has shown how specialized corpus data can be used for the extraction of collocations, terms, and context-specific word senses, which may be used both to enrich the information provided by currently available general dictionaries and to work towards the creation of a truly specialized lexicographic resource. More extensive studies on Italian-English pairs are definitely needed in this perspective; indeed, the parallel corpora foreseen by the LBC project will enable more systematic contrastive analysis that can shed light on both the source and the target language. A possible limitation of this study has to do with the corpus data themselves, which mostly come from ancient texts, especially *Le Vite*. While most terminology introduced by Vasari in his foundational work remains valid at present (and this is the reason why efforts were made by the LBC team to include the full text of both editions in the Italian corpus), more recent and contemporary texts are needed to make the corpus itself – and the investigations based on or driven by it – more balanced. From this perspective, the Italian LBC corpus should (and, as a 'monitor' corpus, will) be updated so as to include more texts, encompassing a wider range of historical periods covered. In the final analysis, this shows how the results of corpus-assisted studies can provide feedback also from the viewpoint of corpus 'maintenance'.

References

Adamska-Sałaciak, A. (2020). Bilingual Learners' Dictionaries in the Lexicographic Landscape. In M. Szczyrbak, & A. Tereszkiewicz (Eds.), *Languages in Contact and Contrast. A festschrift for Professor Elżbieta Mańczak-Wohlfeld on the Occasion of Her 70th Birthday* (pp. 37-47). Krakow, Jagiellonian University Press.

Baker, P., Hardie, A., & McEnery, T. (2006). *A Glossary of Corpus Linguistics*. Edinburgh, Edinburgh University Press.

Bartsch, S., & Evert, S. (2014). Towards a Firthian notion of collocation. *OPAL – Online publizierte Arbeiten zur Linguistik*, *2*, 48-61.

Ballestracci, S. (2023). Da un segno a tanti segni. L'emergere della polisemia del termine *disegno* nelle traduzioni tedesche delle *Vite* di Vasari. In V. Zotti, & M. Turci (Eds.), *Nuove strategie per la traduzione del lessico artistico. Da Giorgio Vasari a un corpus plurilingue dei beni culturali* (pp. 21-38). Firenze, Firenze University Press. https://doi.org/10.36253/979-12-215-0061-5.04

Berti, B., & Pinnavaia, L. (2012). Towards a bilingual Italian-English dictionary of collocations. In R. Facchinetti (Ed.), *English Dictionaries as Cultural Mines* (pp. 201-221). Newcastle upon Tyne, Cambridge Scholar.

Carpi, E., & Pano Alamán, A. (2019). Diseño, Desarrollo y aplicaciones del Corpus LBC sobre el léxico del arte y el patrimonio cultural. *Revista de Lexicografía*, *XXV*, 149-160. https://doi.org/10.17979/rlex.2019.25.0.5992

De Vere, G. (1912-1915). *Lives of the Most Eminent Painters, Sculptors and Architects by Giorgio Vasari*, Vol. I-X. London, Macmillan and Co. Ld. & The Medici Society.

Dima, G. (2012). A terminological approach to dictionary entries. A case study. *PROCEDIA – Social and Behavioral Sciences*, *63*, 93-98. https://doi.org/10.1016/j.sbspro.2012.10.016

Faaß, G. (2018). Lexicography and corpus linguistics. In P.A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography* (pp. 123-137). Abingdon/New York, Routledge.

Flinz, C. (2023). The multifunctional LBC-Corpora: different aims depending on the user. *Lexicographica*, *39*, 191-208. https://doi.org/10.1515/lex-2023-0010

Farina, A., & Billero, R. (2020). Corpus in "natural" language versus "translation" language: LBC corpora, a tool for bilingual lexicographic writing. In D.F. Iezzi, D. Mayaffre, & M. Misuraca (Eds.), *Text Analytics: Advances and Challenges. JADT 2018. Studies in Classification, Data Analysis, and Knowledge Organization.* (pp. 167-178). Cham, Springer. https://doi.org/10.1007/978-3-030-52680-1_14

Farina, A., & Nicolás Martínez, M.C. (2020). La banca dati LBC. In R. Billero, A. Farina, & M.C. Nicolás Martínez (Eds.), *I corpora LBC. Informatica umanistica per il lessico dei beni culturali* (pp. 7-17). Firenze, Firenze University Press. https://doi.org/10.36253/978-88-5518-253-9

Hätty, A., Schlechtweg, D., Dorna, M., & Schulte im Walde, S. (2020). Predicting degrees of technicality in automatic terminology extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 2883-2889). Kerrville, Association for Computational Linguistics.

Iamartino, G. (2019). Recent bilingual English-Italian lexicography: Insights into usage. *Dictionaries: Journal of the Dictionary Society of North America*, *40*(2), 133-154. https://doi.org/10.1353/dic.2019.0017

Luporini, A. (2023). 'Metaphors of art' in the English translations of Giorgio Vasari's *Le Vite*: A software-assisted enquiry. In V. Zotti, & M. Turci (Eds.), *Nuove strategie per la traduzione del lessico artistico. Da Giorgio Vasari a un corpus plurilingue dei beni culturali* (pp. 65-92). Firenze, Firenze University Press. https://doi.org/10.36253/979-12-215-0061-5.06

Maclehose, L.S., & Brown, G.B. (Trans. & Ed.). (1907). *Vasari on Technique, Being the Introduction to the Three Arts of Design, Architecture, Sculpture and Painting, by Giorgio Vasari*. London, J.M. Dent & Co.

Papiol (1909). *Manuale della lingua teatrale*. Milan, Società Editoriale Milanese.

Rey, A. (1995). *Essays on Terminology, Translated and Edited by J.C. Sager*. Amsterdam/Philadelphia, Benjamins.

Rundell, M. (2018). Searching for extended units of meaning – and what to do when you find them. *Lexicography*, *5*, 5-21. https://doi.org/10.1007/s40607-018-0042-1

Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam/Philadelphia, Benjamins.

Zotti, V. (2017). L'integrazione di corpora paralleli di traduzione alla descrizione lessicografica della lingua dell'arte: l'esempio delle traduzioni francesi delle *Vite* di Vasari. In V. Zotti, & A. Pano Alamán (Eds.), *Informatica umanistica. Risorse e strumenti per lo studio del lessico dei beni culturali* (pp. 105-134). Firenze, Firenze University Press.

Notes

Note 1. E.g., for English, the *Oxford Dictionary of Art*, the *Oxford Concise Dictionary of Art Terms*, or the Getty Research Institute *Art and Architecture Thesaurus Online* (https://www.getty.edu/research/tools/vocabularies/aat/index.html); for Italian, the *Dizionario Enciclopedico dell'Arte* Mondadori, the *Dizionario Arte* Jaca Book, or the *Dizionario dell'Arte* Baldini Castoldi Dalai, which is, in fact, translated from the *Oxford Dictionary of Art*. With reference to English, an up-to-date list of monolingual

resources is maintained by the Berkeley Library of the University of California (https://guides.lib.berkeley.edu/c.php?g=478634&p=3273764).

Note 2. Some bilingual technical dictionaries (e.g., the *McGraw-Hill* published by Zanichelli, or the *Nuovo Marolli* and the *Dizionario tecnico dell'edilizia e dell'architettura* published by Hoepli) also include terms related to the artistic domain, especially to the sub-field of architecture, but are not specialized in this sense. Margherita Palli's *Dizionario teatrale* (Quodlibet, 2021) provides translation equivalents in multiple languages of terms related to the sub-field of theatre (architecture and performance).

Note 3. The LBC corpora can be accessed from https://corpora.lessicobeniculturali.net/. For a detailed overview of corpus construction criteria, see Farina & Billero (2020); Flinz (2023).

Note 4. https://corpora.lessicobeniculturali.net/it/

Note 5. There are two editions of Giorgio Vasari's text: the first was published in 1550 as *Le Vite de' più Eccellenti Architetti, Pittori et Scultori Italiani*; the second, revised and extended edition was published in 1568, with the title *Le vite dei più eccellenti Pittori, Scultori, e Architettori*. The two editions are also referred to as 'Torrentiniana' and 'Giuntina', respectively, after the names of the typographers that first printed and circulated them. Both editions are included in the Italian LBC corpus and make up most of its texts. Descriptive statistics concerning corpus composition can be found at the URL provided above.

Note 6. The appropriate tags to be used in regular expressions depend on the tagset used for the corpus, which can be found under 'corpus information' in Sketch Engine and NoSketch Engine.

Note 7. LogDice is the preferred statistic measure for computing collocations on SketchEngine, for several reasons: see https://www.sketchengine.eu/my_keywords/logdice/.

Note 8. The LogDice measure usually helps to avoid this issue, but it was not as effective as expected in this case, possibly because of the still limited size of the corpus we are considering (LogDice performs well on very big corpora), or because of underlying inaccuracies in lemmatization and part-of-speech tagging of the corpus itself.

Note 9. The two dictionaries can be consulted at the following URLs: https://www.collinsdictionary.com/dictionary/italian-english; https://dictionary.cambridge.org/it/dizionario/. The Cambridge Ita-En Dictionary is in fact based on the Cambridge corpus and on other two dictionaries, Global and Password by Kernerman Publishing, which are best defined as *semi-bilingual* learners' dictionaries (more details on the distinction, which is not deemed relevant for the purposes of this study, can be found in Adamska-Sałaciak, 2020, p. 43).

Note 10. For both the Oxford Dictionary and *Il Ragazzini*, digital editions, access was granted by the University of Bologna.

Note 11. The deverbal noun *disegno* comes from *disegnare*, literally 'to represent something by drawing lines on a surface' (*L'etimologico. Vocabolario della lingua italiana*, Le Monnier).

Note 12. According to the etymological dictionary, Italian *figura* is a borrowing from Latin *figūra* – 'image', 'effigy', 'statue', 'appearance' – in turn derived from *fingĕre*, literally 'to mould', 'to craft' (*L'etimologico. Vocabolario della lingua italiana*, Le Monnier).

Note 13. Cf. the entry for *opera* in the *Vocabolario Treccani*, which mentions *opera d'arte* (En. 'artwork') as only one of the possible, more concrete meanings of the word, together with several others, belonging to different fields: e.g., *opera letteraria* (En. 'literary work'), *opera in musica* (En. 'opera'), but also *opera dell'ingegno* (En. 'intellectual work'), *opera viva* (En. 'quickwork' – in shipbuilding, the bottom part of a ship). Indeed, It. *opera* comes from Latin *ŏpĕra*, plural form of *opus*, *opĕris*, which simply means 'product of work' (see https://www.treccani.it/vocabolario/opera/).

Note 14. The collocate list for *opera* can be found here: https://shorturl.at/gqzQT.

Note 15. Indeed, the *Vocabolario Treccani* defines *boccadopera* without establishing links with *boccascena* (see: https://www.treccani.it/vocabolario/boccadopera/).

Note 16. The collocate list for *figura* can be found here: https://shorturl.at/isDEG.

Note 17. The collocate list for *disegno* can be found here: https://shorturl.at/vLWY3.

Copyrights

Splitting Hairs: Designing Corpus-informed Hairstyling Glossary Entries for a Diverse Audience

Sara Gesuato[1], Erik Castello[1], Viviana Gaballo[1]

[1]Department of Linguistic and Literary Studies, University of Padua, Padua, Italy (Note 1)

Correspondence: Sara Gesuato, University of Padua, Department of Linguistic and Literary Studies, via Vendramini 13, 35137 Padova, Italy. Email: sara.gesuato@unipd.it

Abstract

This paper illustrates a bilingual corpus-driven term extraction and description procedure (Bowker & Pearson, 2002) in the popular field of hairdressing. We present examples of term entries (*fringe, texture, to wear*) derived from ad-hoc corpora compiled on the SketchEngine platform (about 200,000 words total), focusing on translation equivalents and phraseologies. We also consider the advantages of such lexical explorations vis-à-vis the information available in professional publications. We finally present a proposal for the identification and design of terminological entries in a multimodal glossary, which should be comprehensive in content and adaptable to the needs of native and non-native speakers, professionals and laypeople.

Keywords: LSP, glossary, hairstyling, lexicography, lexicology, corpus-driven approach

1. Introduction

As is well-known, language for specific purposes (LSP) is language relevant to a circumscribed domain, and therefore shared by members of a community of practice. These may conceptualize their shared domain as a professional, intellectual and/or emotional field of interest. What a given domain (e.g. business) actually represents to a "stakeholder" (e.g. work to entrepreneurs, interest to stock traders, knowledge to lecturers) depends on how people relate to it (i.e. why it matters to them, how they take part in it, and what they know about it). This will affect their level of familiarity with its communicative conventions.

When considering the *communicative* dimension of a domain, participants can be roughly divided into insiders and outsiders. The former are (future) expert "stakeholders", who do or will know about, know how to act in, and know how to communicate in the domain because of their present or future involvement in it, such as practitioners, trainees, enthusiasts, subject teachers and subject students. The latter include language teachers and students, who need to familiarize themselves with both the subject and the interactional practices of the field.

One distinctive feature of the communicative practices of a specific domain is its lexicon. This includes terms (i.e. technical words), common words with domain-specific senses, and general words (Pearson, 1998, p. 17; cf. Chromá, 2004, p. 15). Terms are assigned clearly defined meanings (Pearson, 1998, p. 9), that is, authoritative definitions identifying agreed-upon pairings of labels and concepts (Rondeau, 1984, p. 19); they are thus used for accurate, unambiguous communication, and have congruent equivalents in other languages (e.g. *recessive [gene]*; *plasma*; *aetiology* in medicine). Also, some common words may take on specialized meanings when used in a given domain, for example *significant* ('important') *achievement* vs *significant* ('not due to chance') *difference* in statistics, and *bug* ('insect') vs *bug* ('error, fault, malfunctioning') in computing. Such words are likely to have only partial equivalents in other languages; they are also potentially ambiguous because their domain-specific meanings emerge from use, and are thus in a state of flux. For example, the expression *flexible shoes* is used "loosely" instead of the term *stitchdown shoes* to refer to shoes which are made by stitching the shoe upper directly to the sole, and which, as a result, are flexible (Blecha, 2009, p. 53). Finally, general words are part of everyday communication, and denote the same broad and familiar notions across domains (Sager, 1990, p. 19), for example, the general word *blood* vs. the term *plasma*, and the phrase *the analysis of causes* vs. the term *aetiology*. These words are likely to have congruent equivalents in other languages. Of the three types of lexicon that characterize a domain, it is common words with specialized meanings that are potentially hard to pin down because their meanings are, literally, not defined, that is, they do not have clear-cut boundaries; rather, they are flexibly adapted to users' communicative needs.

As novices and non-experts, language learners and trainee professionals have to build their knowledge of domain-relevant lexis, both technical and common words, through dedicated resources such as glossaries. These should ideally meet two requirements (Szemińska, 2014): exhaustiveness and user-friendliness, that is, all potentially useful information should be provided, and it should be easily retrievable.

Complete information about the usage of terms fosters subject-field understanding, correct term choice, and idiomatic expression (Bowker, 1998). Therefore, it should not only include definitions of the meanings of words, in a range of contexts (Bowker, 2011), but also corpus-based examples (Condamines et al., 2013), collocations, and indications of semantic relationships with other words (Josselin-Leray & Roberts, 2007). It should also offer descriptions of prototypical use, for example, frequency of use, relevance to genres; Szemińska, 2014), as well as type and degree of congruence with candidate equivalents across languages (Szemińska, 2014).

Such contextually rich information about a word (technically called *knowledge-rich context*; Marshman, 2014, p. 583) may be of help to users if it is organized in an easy-to-read format so that diverse users (e.g. translators, language teachers, language learners), as well as the same user at different stages of their work (e.g. a translator who is making sense of a source text vs checking a target text for accuracy; Szemińska, 2014, p. 119), may effortlessly identify and select the information relevant to them.

To illustrate how LSP terminological needs may be addressed, we will present our analysis of a few technical and common words in the popular field (Gasiglia, 2012) of hairdressing. Relying mostly on corpus data, we aim to describe the context of use of these words, presenting a proposal for glossary entries catering to diverse users, namely (trainee) language practitioners and (trainee) professionals alike. There are two reasons for this choice. One is that, as a professional, but not principally scientific, field (unlike the neighboring field of trichology), hairdressing is of potential interest to the general public and experts alike, and may thus be characterized by the presence of common words taking on domain-specific meanings. The other reason is that, although domain-specific linguistic resources (e.g. manuals, glossaries) on hairdressing do exist, they cater more to the needs of future professionals than future language experts: definitions of terms, explanations of techniques and pictures of hairdos are provided, but information about the context of use of terms (e.g. word collocations), especially useful to language learners, is mostly lacking (Note 2).

In the rest of our paper, we will first overview recent literature on the goals and requirements of LSP glossary-making (Section 2); then we will outline our research method (Section 3); next, we will describe the context of use of select hairstyling words (Section 4); we will conclude with the outline of an ideal glossary entry and suggestions for future perspectives (Section 5).

2. Theoretical framework

This work lies at the intersection of lexicography and lexicology. While lexicography is concerned with the compilation and editing of dictionaries, lexicology is more theoretical, and delves into the principles underlying the organization and functioning of lexical units (Halliday, 2004, p. 3).

Glossary entries are the tangible outcomes of the combined concerns of lexicography and lexicology, as they represent a fusion of practical lexicographic methodologies and theoretical lexicological principles (Halliday, 2004, pp. 7-11). As a matter of fact, lexicography and lexicology collaborate throughout the process of glossary creation as lexicological insights inform lexicographic decisions, and lexicographic practices ensure that theoretical principles are applied practically. This process is often iterative, with lexicographers refining entries based on user feedback (Gaballo, 2013, pp. 126-127), changes in language use (Yallop, 2004, pp. 41-44), and advancements in the field (Halliday, 2004, pp. 16-20). Lexicology therefore provides the theoretical framework, and lexicography adapts it to meet the practical needs of users.

Glossary entries, influenced by lexicological principles, strive to capture the accurate and nuanced meanings of terms within the context of a particular field, ensuring precision in communication (Sager, 1990). Lexicography emphasizes the needs of users ensuring that the information is presented in a way that is accessible and useful for the target audience, whether practitioners, learners, or translators. Glossary entries, informed by lexicographical practices, are contextually relevant, often including information about how terms are applied within specific scenarios or industries (*Dictionary of Lexicography*, 1998). They may include cross-references to related terms, so that they create a network of interconnected entries that enhances the user's understanding of the terminology within a broader context (Landau, 1989).

What glossaries share with dictionaries is first of all their language orientation, i.e., both of them aim to provide explanations, definitions, and understanding of terms within a specific domain, and both of them serve as reference materials, aiding users in comprehending the meaning and usage of words or terms. Their scope and purpose, however, may differ: dictionaries are comprehensive tools that cover a wide range of words, including their meanings, pronunciations, etymologies, and usage; while glossaries are focused on specific subjects, industries, or contexts, and provide concise explanations for terms related to a particular field (Gaballo, 2022). Dictionaries and glossaries also share common structural elements such as headwords, definitions/explanations, and sometimes examples, and the creation of both involves research, compilation, and curation of relevant terms and their meanings. On the other hand, they differ in terms of depth of information and target audience: dictionaries provide in-depth information on word usage, multiple meanings, and historical context, while glossaries offer concise

explanations, emphasizing clarity over depth; dictionaries are aimed at a general audience, including native speakers and language learners, while glossaries are tailored for specific audiences within particular domains.

Dictionaries are invaluable tools in language learning and education – albeit not the only ones – and provide learners with a wealth of information on vocabulary, grammar, and idiomatic expressions. In literary analysis and interpretation, dictionaries aid readers and scholars in understanding the nuances of words, ensuring a more profound exploration of texts. They play a crucial role in general communication and writing, helping individuals choose precise and appropriate words to convey their thoughts accurately (Béjoint, 2000).

Glossaries, on the other hand, specialize in industry-specific terminology, offering concise explanations of terms relevant to particular fields or sectors. They are indispensable in technical communication and documentation, providing a quick reference for technical terms and jargon used in manuals, guides, and reports. Glossaries are essential for field-specific training and education, as they assist learners in grasping the specialized vocabulary and concepts unique to a particular discipline. In translation and interpretation within specialized domains, glossaries prove invaluable as they ensure accuracy and consistency in rendering technical terms across languages.

Zooming in from this general comparison between dictionaries and glossaries, and giving a cursory glance at a variety of field-related dictionaries and glossaries (see references for a list), we can spell out the specific needs, requirements and characteristics of glossary entries meant for either the technical-scientific domain or a laypeople context, and for either trainee practitioners or trainee translators/mediators/LSP language specialists, as specified in Table 1.

Table 1. Contrasting Needs/Requirements/Characteristics of dictionary vs glossary entries

Technical-Scientific Domain: Trichology	*Laypeople Context: Hairdressing*
Dictionary Entries: Detailed scientific explanations, molecular structures, medical contexts.	Dictionary Entries: Everyday language, common terms in haircare.
Glossary Entries: Specialized terms related to hair and scalp health, treatments, and research.	Glossary Entries: Practical terms related to styling, cutting, and salon procedures.
Trainee Practitioners	*Trainee Translators/Mediators/LSP Language Specialists*
Dictionary Entries: Emphasis on practical application and execution.	Dictionary Entries: Extensive language knowledge, nuances, idiomatic expressions.
Glossary Entries: Industry-specific terms, best practices, and case-specific scenarios.	Glossary Entries: Precision in translating specialized terms, cultural considerations, and context-specific meanings.

The next section of this paper will outline the method adopted to extract and organize corpus-based information about LSP terms useful to cater for both practitioners and trainee translators/language learners.

3. Methodological framework

Research in the fields of terminology and lexicography has recently greatly benefited from advances in corpus linguistic methodologies (Condamines et al., 2013, p. 583; Bowker, 2018, pp. 143-144). We therefore decided to carry out a corpus-informed investigation of hairdressing terminology. Specifically, the aim of our project was to single out relevant terms and set up glossary entries for L2 trainee professionals and Italian students of English language and translation.

In order to identify candidate terms, we looked at the language currently used in (semi-)professional communication, and to study it using corpus-based methods. We availed ourselves of the online platform *SketchEngine* to construct a specialized English corpus, extract terms from it and analyze them. We provided seed words to the software (*hairdo, hairstyle, haircut, hair, hairstylist, hairdresser*), and kept the default settings with

the exception of "Sites list", which we limited to ".uk" websites. The system downloaded 126 relevant texts from the Internet, cleaned them up and POS-tagged them, and we called this the *UK-Hairstyle* corpus (142,869 tokens; 14,262 types). As can be seen in Appendix 1, most of the texts were retrieved from online British magazine articles or websites devoted to hairstyling. We followed the same procedure to compile a comparable Italian corpus. We input the seed words/phrases *pettinatura, acconciatura, taglio di capelli, capelli, acconciatore* and *parrucchiere*, in response to which the system collected 186 relevant texts into a corpus we called *IT-Acconciatura* (225,505 tokens; 17,829 types). Appendix 2 shows that the Italian texts, too, mainly come from magazine articles and websites on hairstyling.

To examine the corpora, we then used the in-built tools "Wordlist (Frequency list)", "Keywords (Terminology extraction)", "Concordance (Examples of use in context)", and "Word Sketch (Collocations and word combinations)". We obtained the single-word keyword list of the *UK-Hairstyle* corpus by comparing it against the *English Web 2021* corpus. Table 2 shows the keywords that we agreed could represent good candidate terms for analysis, listed in decreasing order of key score. It is worth noting that the keywords in bold rank rather low in terms of keyness as they are polysemous words used in other semantic fields as well (i.e. *bun, texture, bob, bang, length*).

Table 2. Candidate terms extracted from the *UK-Hairstyle* corpus through the *Sketch Engine* Keyword list tool. The items in bold are polysemous and may be used in fields other than hairdressing.

Terms from KWL	Freq.	Key score	cont.	Freq.	Key score
balayage	65	429.18	texturising	11	77.21
blow-dry	65	401.96	side-swept	11	76.53
ponytail	78	285.76	bouncy	24	72.69
pixie	75	215.90	cornrows	11	72.62
shag	56	211.35	**bun**	**39**	**59.37**
mullet	47	179.90	beehive	15	54.63
updo	27	164.09	wig	36	54.03
fringe	190	146.02	chignon	8	53.84
plait	30	143.33	perm	14	53.83
parting	46	142.39	brunette	21	53.45
curl	172	125.73	bouffant	8	53.20
lob	43	125.03	hairline	13	52.57
up-do	18	121.67	**texture**	**166**	**51.13**
choppy	34	116.29	half-up	7	49.03
wavy	39	111.43	buzzcut	7	48.79
frizz	21	111.42	air-dry	8	48.62
tousle	19	108.88	swept-back	7	48.49
straightener	20	107.39	quiff	7	48.05
braid	62	106.66	nape	10	47.96
mixie	12	83.14	**bob**	**266**	**47.80**
blonde	102	80.11	**bang**	**74**	**36.83**
			length	**212**	**20.30**

We were able to identify some more candidate terms by looking at recurrent words in the *UK-Hairstyle* corpus Wordlist. The words listed in Table 3 are arguably frequent in general English, and are used in other semantic fields besides hairdressing, which explains their absence from the keyword list.

Table 3. Candidate terms identified in the *UK-Hairstyle* corpus through the *Sketch Engine* Wordlist tool.

Terms from WL	*Freq.*
end	109
even	105
extension	68
full	87
highlight	83
root	66
shape	139
straight	75
volume	123
wave	116
wear	115

We also consulted the relevant literature on the creation of glossary entries, and searched the web for online resources about hairdressing. We retrieved a series of online glossaries and manuals for trainee and professional hairstylists in English and in Italian, with a view to cross-checking and confirming the corpus-driven selection of the most recurring terms used in the profession (see the list at the end of the paper). However, we noticed that such materials often list sets of terms that do not fully match, and provide partial or even contradictory information about them.

We used the Concordance and Word Sketch tools to obtain corpus data from the *UK-Hairstyle* corpus in the form of concordance lines, collocates and phraseological patterns. With this data we prepared drafts of glossary entries, which we then repeatedly revised by comparing and integrating them with information from dictionaries, online glossaries and manuals. We also derived corpus data from the *IT-Acconciatura* corpus, which we explored in search for likely Italian translation equivalents and phraseological patterns.

This overview of the methodological framework has paved the way to the next section devoted to the descriptions of the select terms and their contexts of use.

4. Findings and discussion

For the present study, we decided to focus on three words. The first is a word scoring a high keyness value, which is very specific to the language of hairdressing: *fringe*. We then opted for a word with a low key value, which is extensively employed in other sectors too: *texture*. Finally, we chose a word that is used in many expressions in the field of hairdressing as well as in general English, and represents a different part of speech: the verb *to wear*.

4.1. Fringe(s) and bangs

In the context of British English[(BrE)] hairstyling, a *fringe* (pronunciation /frɪndʒ/) refers to a section of hair that is cut and styled to fall over the forehead, framing the face (see example 1). Fringes may vary in lengths and styles, and hairstylists may incorporate them to add texture, enhance facial features, or create a specific look for their clients. Fringes can be straight across, side-swept, layered, or textured, depending on the desired aesthetic and the individual's face shape.

(1) […] From bobs and lobs to braids and full fringes, there are so many amazing hairstyles out there.[(BrE)] (*UK-Hairstyle* corpus, 2023)

This hairstyle is also commonly known as *bangs* in American English.[(AmE)] (see example 2).

(2) […] To put it frankly, Charlize Theron (and her fierce side-swept bangs) aren't f–ing around in the new trailer for The Old Guard.[(AmE)] (English Web Corpus *enTenTen21*)

Fringe is a countable noun, while *bangs* is only used in the plural (Note 3). Both terms encompass various styles of front-facing haircuts, ranging from straight-across fringes[(BrE)]/bangs[(AmE)] to side-swept or

textured[(BrE)]/layered[(AmE)] variations, allowing for a range of looks tailored[(BrE)]/customized[(AmE)] to individual preferences (see examples 3 and 4).

(3) […] Are you going for a fringe that ends at your eyebrows?[(BrE)] (*UK-Hairstyle* corpus, 2023)

(4) […] There are different types of bangs, curls and other hair dos that you can try. (English Web Corpus *enTenTen21*)

As a verb, *to fringe*[(BrE)] or *to bang*[(AmE)] is used to describe the action of cutting or styling the hair at the front to create the desired fringe[(BrE)]/bangs[(AmE)] effect (see examples 5 and 6); however, occurrences of this type of use are rare.

(5) […] To fringe or not to fringe… that is the question! Over time, we've all faced many a fringe-hair dilemma; wondering what fringe hairstyle to try and whether or not it will suit us.[(BrE)] (Allthingshair.com > Fringe hairstyles)

(6) She decided to bang her hair, opting for a trendy and youthful appearance.[(AmE)]

The term can also be used as an adjective to describe hairstyles that feature a prominent front section of hair falling over the forehead (see examples 7 and 8).

(7) […] Getting a fringe cut in can be nerve-wracking, especially if your waves look different from day-to-day.[(BrE)] (*UK-Hairstyle* corpus, 2023)

(8) […] The side-by-side photo montage she made of my bangs haircut with an upside down bowl of Japanese rice kills me every time.[(AmE)]. (*UK-Hairstyle* corpus, 2023)

The humorous definition of *fringe* as 'politically correct hair terminology for *bangs*' underlines the suggestive connotation of the polysemic term *bang(s)* when in the context of a private relationship. Since the word *bang* is perceived as offensive in some cultures, the more neutral *fringe* appears in its stead.

Worth-noting is the derivational morpheme *fringe-esque*, which is often used to describe things that exhibit features typically associated with fringes, such as fashion, design, or artistic elements, as in the more exotic *Jane Birkin-esque*. *Fringe-esque* carries positive connotations, often associated with creativity, individuality, and a distinctive aesthetic. It suggests a certain level of uniqueness and originality in the context to which it is applied, as shown in the following examples:

(9) If you don't have the 'ideal' four finger forehead you can still have a fringe-esque hairstyle. (Beaut.ie)

(10) Caesar Haircut. A men's buzz cut with shorter proportions on the sides and the back and shorter fringe-esque styling at the front of the head; originated from Julius Caesar, who wore his hair in this manner. (thetrendspotter.net)

Both *fringe* and *bangs* exhibit common collocates such as *straight*, *curly*, and *side-swept*, reflecting different styles. They both share colligational structure that include adjectives describing the characteristics of the hairstyle (e.g., *brow-grazing, eyelash-skimming, flicked-out*), verbs associated with hairstyling actions (e.g., *cut*, *trim*, *curl*), and prepositions indicating the location on the head (e.g., *fringe on the forehead, bangs across the face*). Occasionally, both *fringe* and *bangs* occur in the same phrase/sentence (e.g., *Long bangs can be combined with fringe haircuts*).

The Italian equivalent of *fringe/bangs* is *frangia*, more commonly used in its variant *frangetta*. The collocations and colligations of *frangia* in the Italian hairdressing industry reflect the variety of styles, cuts and aesthetic approaches associated with this distinctive element of hair fashion.

Collocations of *frangia* in the hairstyle industry are varied and depend on the style, length and care of the fringe (e.g. *frangia dritta, lunga, corta, laterale, sfilata, spettinata, retro, moderna, classica*). Colligations reflect the cut, style and daily management of the fringe, from both the customer's and the hairdressing professional's perspectives (e.g., *taglio con frangia, frangia ben curata, portare la frangia lateralmente, tagliare/stilizzare/sfoltire/pettinare la frangia*). The use of *frangia* in prepositional contexts highlights the position of the fringe on the face, suggesting its impact on the shape and overall aesthetic effect? (e.g., *frangia sopra gli occhi*, *frangia intorno al viso*). Professional colligations show how the fringe is a key element in the offer of services in beauty salons and in stylistic consultancy (e.g., *salone di bellezza con specializzazione in frange, consulenza stilistica sulla frangia, prodotti per la cura della frangia*).

The collocations and colligations of *frangia* in the hairdressing industry reflect the variety of styles, cuts and aesthetic approaches associated with this distinctive element of hair fashion (e.g., *frange sfilate dal design unico, accessori che accentuano la frangia, stile boho con frangia, frange che seguono le ultime tendenze di design*).

4.2. Texture

The word *texture* (pronunciation /ˈtɛkstʃə(r)/) is a noun used in general English to indicate the way a surface, substance or cloth feels when one touches or tastes it. It can be used as an uncountable or countable noun (*textures*). If used in the domain of *hair,* it refers to the natural pattern or structure of an individual's hair strands. The technical

definitions found in available hairstyling glossaries range from "the circumference of your hair", "the type of hair: fine, medium or thick", "hair that has texture in it is hair that is not all the same length" to "dimension added to the hair by cutting it vertically". A glossary entry should acknowledge the diverse specific meanings and denotations that the term might take according to the specific context of use.

The concept of *texture* can be explained as the "thickness of your hair", yet a distinction should be drawn between the use in everyday professional registers and that in technical registers. In the former contexts, the term tends to indicate the thickness of one's whole hair often resulting from cutting or treating the hair, while in the latter it refers to the natural thickness (diameter) of individual hair. The hyponyms *fine, medium* and *thick/course texture* are often used to classify the dimension of each single hair. A concept related to *texture* is *hair density*, which specifically describes the amount of hair in a given area of the scalp, and can be classified into *thin, medium,* and *thick density.*

Data from the *UK-Hairstyle* corpus reveals that *texture* is often preceded by the noun *hair* and that a large number of verbs and adjectives collocate with it on its left. The most notable collocating verbs include *add, boost, create, deliver, enhance, forgo, give, lay, maintain, restore, soften, suit, tousle, undo,* and *wear,* while the adjectives are *added, beachy, bouncy, curly, faux, feathery, intricate, lightweight, messy, natural, soft, structured, rough, thick, undone,* and *wavy.* Most of these collocates suggest that the type of *texture* meant in the texts is of the former type, that is it describes what the hair looks like after some type of treatment. The nouns that combine with *texture* similarly tend to denote characteristics of hair obtained after some treatment (e.g., *blend*, *body, type, layer, length, movement, shade, shape, shine, volume* and *weight).* Some of these might also be used as near-synonyms of *texture* (e.g., *volume, body).* Finally, the phrases in which *texture* occurs show that the nouns may be accompanied by varied quantifiers: *a little, tons of, a bit of, full of, plenty of* (when used as an uncountable noun), and *a range of* (*"textures",* when used as a countable noun).

The collocates just described could be shown to glossary users, while the following extracts could be proposed in the *Examples* section as illustrations of possible contexts of use of the term:

(11) Plus, it's a great option for thin hair, as layers help to add movement and texture to naturally flat styles. (*UK-Hairstyle* corpus, 2023)

(12) Keeping it in its natural texture from time to time also gives the hair a break from heated tools which can damage it in the long run. (*UK-Hairstyle* corpus, 2023)

(13) If your hair is fine and you struggle to get anything from it, this air-drying cream will boost texture and deliver a little bit of volume. (*UK-Hairstyle* corpus, 2023)

(14) Her feathered hair had texture, volume and movement. (*UK-Hairstyle* corpus, 2023)

(15) If your hair is naturally thin and straight, let your hair dry naturally and embrace its natural shape and texture. (*UK-Hairstyle* corpus, 2023)

(16) Clarke and his team cut hair dry to enhance your hair's natural movement, taking into consideration its weight, texture and condition. (*UK-Hairstyle* corpus, 2023)

(17) It works beautifully with a range of textures, with delicate curly-girl, wavy or straight fringes. (*UK-Hairstyle* corpus, 2023)

The glossary entry should also list and exemplify the words that are morphologically related to *texture*: *textures* (13), *textured* (21), *texturising* (11), *texturizing* (1), *texturiser* (1), *texturing* (1).

Online sources suggest *consistenza* (*sottile* and *grossa*), *consistenza del fusto capillare* and the Anglicism *texture* as Italian specialized equivalents of *texture.* Data from the *IT-Acconciatura* corpus not only confirms that *consistenza* (31), *corpo* (33) and *texture* (27) can denote the same concept(s) as the English term, but also point to *volume* (251 occurrences) as yet another near-synonym. Specifically, the verb collocates shared by *volume, corpo* and *consistenza* are *aggiungere, dare* and *conferire.* Other verbs that are used before *volume* are *aumentare, creare, conferire, ridurre, perdere, donare, riempire, contenere, ottenere,* and *apportare.*

The occurrence of the Italian equivalents in combination with some of these collocates is exemplified by the following extracts from the *IT-Acconciatura* corpus:

(18) Un dettaglio che permette di dare volume e movimento alla chioma senza troppa fatica. (*IT-Acconciatura* corpus, 2023)

(19) Le texture, i volumi e le linee si uniscono ad infinite combinazioni di colore. (*IT-Acconciatura* corpus, 2023)

(20) […] uno stile che si adatti alla consistenza del fusto. (*IT-Acconciatura* corpus, 2023)

(21) […] un ottimo modo di mettere in risalto la texture della chioma. (*IT-Acconciatura* corpus, 2023)

(22) […] lucidare, lisciare, donare corpo e volume. (*IT-Acconciatura* corpus, 2023)

These examples could be made available to users as a component of the glossary entry.

4.3. To wear

The verb *to wear* (pronunciation: BrE /wɛə(r)/, AmEn /weər/) is a common word. As a transitive verb, its basic meaning is 'to have clothes, shoes, jewels or accessories on a part of one's body'; by extension, it may also mean 'to show a given emotion on one's face'. When applied to the field of hairdressing, it acquires a more specific meaning, which is 'to have or arrange one's hair or beard in a particular style'. The shared meaning in all these contexts is 'to have and display', which is thus partly adapted in relation to the specific object of wearing.

As the *UK-Hairstyle* corpus data reveal, the transitive verb *to wear* in the active voice is followed by a direct object (i.e., a patient), and then, very often, also by an object complement (i.e., a property of the patient). The direct object slot can be occupied by general nouns denoting hair (e.g., *hair, shag, wig*), nouns denoting parts of hair (e.g., *ponytail*, *braid*, *extension*) and others denoting the look or style of hair (e.g., *cut*, *up-do*, *hairstyle*). The direct object complement slot is usually occupied by adjectives illustrating natural or specially created properties of hair (e.g., *smooth, short, super-high, straight, fluffy, elegant,* and less frequently by particles, prepositional phrases or adverbs indicating how one's hair is worn (e.g., *up, in a ponytail, gracefully,* as shown by the examples below:

(23) Elizabeth wore her hair down. (*UK-Hairstyle* corpus, 2023)

(24) This can get boring, so you may want to try wearing your hair just behind your ears instead. (*UK-Hairstyle* corpus, 2023)

(25) As far back as Roman times, it was popular for women to wear their hair long. (*UK-Hairstyle* corpus, 2023)

(26) Adding curls or waves to your hair will really emphasise the layers which is the key to wearing this trending style. (*UK-Hairstyle* corpus, 2023)

Concordances also show that the verb *to wear* may co-occur with the verb *to grow* in coordinated clauses, or occasionally replaced by the near-synonym *to sport*; e.g.:

(27) These are haircuts that can be worn and grown without too much attention and they stay looking strong with you. (*UK-Hairstyle* corpus, 2023)

(28) Valentino sported his hair high, tight and ultra glossy, whipped up into a perfectly varnished side parting. (*UK-Hairstyle* corpus, 2023)

According to general dictionaries, the translation equivalents of *to wear* in Italian are *portare, indossare, mettersi addosso* and *vestire*. However, the *IT-Acconciatura* corpus data reveals that only *portare*, and much more rarely *sfoggiare*, occur in the same phraseological pattern as *to wear*. That is, *portare* is followed by the translation equivalents of the direct objects of *to wear* (e.g., *caschetto*, *ciocca*, *frangia*, *riga*, *taglio*, *treccia*), and then by adjectival, prepositional or adverbial phrases denoting the properties of hair, as in the following examples:

(29) Per essere autonome al 100%, l'ideale è portare un taglio pari o leggermente scalato. (*IT-Acconciatura* corpus, 2023)

(30) Taglio a caschetto: chiamato anche carré, è un taglio essenziale e intramontabile che si porta a ogni età. (*IT-Acconciatura* corpus, 2023)

(31) Non importa se porti la riga nel mezzo, a sinistra o a destra. (*IT-Acconciatura* corpus, 2023)

(32) Lo chignon portato in alto e diviso in due è giocoso e poco impegnativo. (*IT-Acconciatura* corpus, 2023)

(33) Porta tutti i capelli in avanti. (*IT-Acconciatura* corpus, 2023)

(34) […] un'unica acconciatura da portare di lato o lungo la schiena. (*IT-Acconciatura* corpus, 2023)

(35) […] la treccia morbida, da portare laterale o delicatamente appoggiata sulla schiena. (*IT-Acconciatura* corpus, 2023)

(36) Per valorizzare gli occhi e gli zigomi si può sfoggiare una lineare frangia lunga da portare pettinata aperta su tutta la fronte. (*IT-Acconciatura* corpus, 2023)

The English and the Italian examples show that *to wear* and *portare* illustrate the same phraseological pattern as general dictionaries, but with a wider range of direct objects and object complements. This reveals that these verbs are particularly relevant to the subfields of haircuts and hairstyles. English hairstyling glossaries and manuals do not have entries for *to wear*, but contain instances of *to wear* followed by a direct object, or a direct object and its complements, and also instances of *wearable* and *wearer*, as in the following examples:

(37) […] giving the wearer the appearance of having longer and thicker hair. (*UK-Hairstyle* corpus, 2023)

(38) […] she has the option of wearing a ponytail. (*UK-Hairstyle* corpus, 2023)

(39) […] add a touch of softness to make it more wearable! (*UK-Hairstyle* corpus, 2023)

A hairstyling glossary entry of *to wear* should thus include numerous examples of the verb so as to be of use especially to EFL students and translators who need to be sensitized to the lexico-grammatical scope of the verb, and its comparable lexico-grammatical translation equivalents.

5. Conclusions

As illustrated in this paper with reference to hairdressing, LSP is not a property of technical, scientific and professional domains, but rather of any field of focused interest. Also, it is not restricted to communicative events involving experts, but also quasi-experts and non-experts (Mikkelsen, 1991, p. 100, quoted in Humbley, 2018, p. 319). Finally, it does not include a single genre or text type – the same topic may be dealt with in a variety of contexts ranging from professional/client encounters, through promotional, informative, regulatory communications, to academic interactions.

The lexis of an LSP is necessarily varied and multi-functional, and thus likely to be differently (un)familiar to different interested parties like trainee professionals, language students and language practitioners (Caruso & Ruffolo, 2014, p. 341). To be of use to varied users, lexicographic resources need to be rich in information and flexible in structure. Our sample analyses of select words relevant to hairstyling has shown that it is possible to envisage a composite profile of glossary entries catering to users' varied and variable communicative needs (e.g. ascertaining the meaning of words, determining the scope of their denotation, exploring their combinatorial patterns). At the same time, we also argue that such composite information has to be easy to search and retrieve for efficient use by experts, trainees and non-native speakers.

Glossary entries are the result of a symbiotic relationship between lexicography and lexicology. Lexicological principles provide the theoretical underpinning, while lexicographic practices ensure that the glossary entries meet the practical communication needs of the users within a specific domain or field. The collaboration of these two disciplines contributes to the precision, clarity, and usability of glossary entries. In hairstyling, *fringe*[(BrE)] or *bangs*[(AmE)] is a versatile term that encompasses various front-facing haircuts, offering individuals a wide range of options to express their personal style and stay on-trend. The terms can be used as nouns (countable in BrE, only plural in AmE), verbs, and also as adjectives. The behavior of *fringe* and *bangs* remains largely consistent between British and American English. Both linguistic variations share common collocations and colligations, suggesting a shared understanding and usage of these terms in the context of hairstyling, regardless of the English variant. However, while both *fringe* and *bangs* share some commonalities in terms of collocations and colligations, subtle differences exist in their associated adjectives and prepositions, reflecting nuances in styling and usage within the hairstyling domain. Worthy of note is the derivational morpheme *fringe-esque*, which conveys a sense of uniqueness, style, or a slightly unconventional and decorative quality.

Texture takes on two specific meanings in the hairstyling domain: in technical registers it tends to denote the natural thickness of each hair, while in everyday professional registers it indicates the thickness of one's whole hair. Corpus data reveals that *texture* is often listed alongside other hair characteristics, such as *body*, *movement*, and *volume*. It also reveals a series of verbs (e.g., *boost, create, deliver*), adjective (e.g., *beachy, bouncy, feathery*) and quantifiers (e.g., *tons of, a bit of, a range of*) preceding the term, which might prove useful to trainee professionals. Finally, both corpus data and online sources suggest that *consistenza (sottile e grossa)*, *corpo* and the loanword *texture* are the Italian equivalents of the term.

To wear is a general word, and the sense it activates in the context of hairstyling is a familiar notion. As such, it does not show up as an entry in specialized glossaries, and its phraseologies are only partly illustrated in specialized texts. Corpus data shows not only a much wider range of its co-textual patterns, but also comparable patterns in the L2, both of which are crucial to language students and trainee translators.

The experience gained through this study suggests that preparing glossary entries involves a cyclical workflow process, which is in need of constant monitoring. This includes the collection of online resources and corpus data, the analysis of data in context, hypothesis testing, cross-linguistic and cross-variety comparisons, and the organization of the retrieved information to make it accessible to the target users. The project has shown that the keyness of words is not the only parameter that should be considered to identify candidate terms. The reason is that there are words that score low on keyness, but are relevant: they straddle the specialized language of hairdressing and general English (e.g., *texture* and *to wear*), and their specialized use might be challenging to master for language learners and trainee translators.

We propose that a thorough glossary entry should include many components, and that these should be grouped under a few broad headings. If designed in a digital format, these headings could be presented as a hamburger menu, so that they could be clicked on and expanded as the need arises. This proposal draws on Bowker and Pearson (2002) and Bowker (2023).

Semantic profile
- specification of the sub-field a given word belongs to (e.g., haircuts, hair part/feature, hair treatment/procedure,

equipment)

- technical definition as found in expert sources, if applicable
- non-technical definition as found in non-expert texts, if attested
- related words: near-synonyms, hyponyms, superordinates and their definitions
- general definition in everyday language
- denotational representation: links to pictures/videos

Grammatical profile

-grammatical category (i.e., part of speech)

-pronunciation in IPA and with an audiofile

- attested word forms
- related words (e.g., morphological derivates)

Phraseological profile

- collocations
- coordinated words of the same part of speech
- colligations

Usage

- register: everyday vs technical
- language variety, e.g., UK/US

Examples:

- phrases
- clauses/sentences

Translation into Italian

- from general dictionaries, i.e., outside the specific domain
- from specialized sources
- corpus-driven, based on shared collocates
- examples

Linguistic details:

- etymology
- frequency of occurrence across genres and over time

Each of the above components is potentially relevant to glossary users: precise definitions of technical terms ensure efficient and unambiguous communication among experts. Visual aids clarify to quasi-/non-experts their denotation. Organizing terms according to sub-domains clarifies the ontological space they belong to (cf. Blecha, 2009, p. 44; Pawłowicz, 2014, p. 55). Concordances of candidate equivalents are helpful for accuracy in translating propositional content across languages. This is in line with Levačič's (2006) notion of entry as quoted in Berginc (2014, p. 388), with recent descriptions of the microstructure of LSP lexicographic entries (Humbley, 2018), and with Berginc's (p. 390) own proposal about the digital format of the entry itself. (Note 4) More generally, Jackson (2018) observes how online lexicography has to aim at corpus integration, efficiency of access and customization (p. 540), offering a wide range of search possibilities (p. 543). In particular, accessibility of information that users look up may have to rely on a dynamic data display (Gouws, 2014, p. 164) with a multi-layered structure so that different blocks of information may be searched at will.

References

Béjoint, H. (2000). *Modern Lexicography: An Introduction*. Oxford: Oxford University Press.

Berginc, N. L. (2014). A corpus based e-dictionary of terminology as a body of knowledge. In G. Budin & V. Lušicky (eds.), *Languages for Special Purposes in a Multilingual, Transcultural World, Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria*. Vienna: University of Vienna, 386-392.

Blecha, J. (2009). *Compiling a Glossary of Terminology Used in Shoemaking*. Bachelor thesis, Masaryk University.

Bowker, L. (1998). Using Specialized Monolingual Native-Language Corpora as a Translation Resource: A Pilot Study. *Meta: Journal des traducteurs / Meta: Translators' Journal*, *43*(4), 631-651.

Bowker, L. (2011). Off the record and on the fly: Examining the impact of corpora on terminographic practice in the context of translation. In A. Kruger, K. Wallmach & J. Munday (Eds.): *Corpus-based Translation Studies: Research and Applications*. London/New York: Continuum, 211-236.

Bowker, L. (2018). Lexicography and terminology, in P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography* (pp. 138-151). London/New York: Routledge.

Bowker, L. (2023). *De-mystifying Translation: Introducing Translation to Non-translators*. London/New York: Routledge.

Bowker, L., & Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*, London/New York: Routledge.

Canazza, G. et al. (2020). *On Hair. English for Hairdressers*. Torino: Edisco.

Caruso, A., & Ruffolo I. (2014). ESP self-compiled corpora in an Italian academic setting. In G. Budin & V. Lušicky (Eds.), *Languages for Special Purposes in a Multilingual, Transcultural World, Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria* (pp. 340-351). Vienna: University of Vienna.

Cascella, V. (2012). *Hairstyle. English for Hairdressers. Tutto il Sapere della Grammatica e della Lingua Professionale*. Bergamo Pontenarica: Editrice San Marco.

Chromá, M. (2004). *Legal Translation and the Dictionary.* Tübingen: Max Niemeyer Verlag.

Condamines, A. et al. (2013). Using Comparable Corpora to Characterize Knowledge-Rich Contexts for Various Kinds of Users: Preliminary Steps, *Procedia - Social and Behavioral Sciences*, *95*, 581-586.

Dictionary of Lexicography (1998). London/New York: Routledge.

Gaballo, V. (2013). A touch of color: A linguistic analysis of the use of color terms in the language of tourism. In D. Höhmann (Ed.), *Tourismuskommunikation: Im Spannungsfeld von Sprach- und Kulturkontakt* (pp. 119-140). Bern: Peter Lang

Gaballo, V. (2022). Cross-linguistic and cross-cultural conceptualization of specialized terms in corporate culture. In B. Lewandowska-Tomaszczyk & M. Trojszczak (Eds.), *Language Use, Education, and Professional Contexts* (pp. 21-40). Cham: Springer.

Gasiglia, N. (2012). Exploiter les énoncés de corpus thématiques à haut rendement pour décrire des prédicats verbaux de lexiques spécialisés de masse", *Lexicographica*, *28* ('Corpora and Lexicography'), 207-232.

Gouws, R. H. (2014). Article structures: Moving from printed to e-dictionaries, *Lexicos*, *24*, 155-177.

Halliday, M.A.K. (2004). Lexicology. In M.A.K. Halliday, W. Teubert, C. Yallop & A. Čermáková, *Lexicology and Corpus Linguistics* (pp. 1-22). London and New York: Continuum.

Humbley, J. (2018). Specialised dictionaries. In P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography* (pp. 317-334). London/New York: Routledge.

Jackson, H. (2018). English lexicography in the Internet era. In P. A. Fuertes-Olivera (Ed.), *The Routledge Handbook of Lexicography* (pp. 540-563). London/New York: Routledge.

Josselin-Leray, A. & Roberts, R.P. (2007). La définition des termes dans les dictionnaires généraux unilingues: Analyse de quelques exemples du domaine de la volcanologie à la lumière d'un corpus de vulgarisation. In L'Homme, M. C. & Vandaele, S. (Eds.), *Lexicographie et Terminologie : Compatibilité des Modèles et des Méthodes* (pp. 141-188). Ottawa: Presses de l'Université d'Ottawa.

Landau, S. I. (1989). (2nd edn). *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.

Marshman, E. (2014). Enriching Terminology Resources with Knowledge-rich Context: A Case Study. *Terminology, 20*(2), 225-249.

Pawłowicz, E. (2014). Term selection methodology for the study of an underresearched subject field: The case of equestrian specialized language. In G. Budin & V. Lušicky (Eds.), *Languages for Special Purposes in a Multilingual, Transcultural World, Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria* (pp. 55-73). Vienna: University of Vienna.

Pearson, J. (1998). *Terms in Context*. Amsterdam/Philadelphia: John Benjamins P.C.

Rondeau, G. (1984). *Introduction á la Terminologie*. Québec: Gaëtan Morin.

Sager, J. (1990). *A Practical Course in Terminology Processing*. Amsterdam/Philadelphia: John Benjamins P.C.

Szemińska, W. (2014). Translating law into dictionaries, or why one dictionary is not enough. In G. Budin & V. Lušicky (Eds.), *Languages for Special Purposes in a Multilingual, Transcultural World, Proceedings of the 19th European Symposium on Languages for Special Purposes, 8-10 July 2013, Vienna, Austria* (pp. 118-125). Vienna: University of Vienna.

Yallop, C. (2004). Words and meaning. In M.A.K. Halliday, W. Teubert, C. Yallop & A. Čermáková (Eds.), *Lexicology and Corpus Linguistics* (pp. 23-71). London and New York: Continuum.

List of online resources

▪ https://nutrafol.com/blog/hair-texture-hair-types/ (last visited on 19/12/2023)

▪ https://www.wisteriaavenue.co.uk/a-z-hairdressing-terms-explained/ (last visited on 19/12/2023)

▪ https://lush-hair-folk.com/hair-salon-glossary-terms/ (last visited on 19/12/2023)

▪ https://www.thetrendspotter.net/key-hairstyle-haircut-terminology-guide/ (last visited on 19/12/2023)

▪ https://nutrafol.com/blog/hair-texture-hair-types/ (last visited on 19/12/2023)

▪ https://secretparrucchieri.it/wp-content/uploads/2018/07/5_regole_per_essere_bellissima_2018.pdf (last visited on 19/12/2023)

▪ https://www.taglicapelliricci.it/capelli-ricci-grossi-fini-o-ispidi-questione-di-texture/ (last visited on 19/12/2023)

▪ Cambridge Online Dictionary (https://dictionary.cambridge.org/)

▪ Longman Online Dictionary (https://www.ldoceonline.com/dictionary)

Notes

Note 1. The authors are jointly responsible for the design of the study and the writing of Section 5. Additionally, Sara Gesuato wrote Sections 1 and 4.3., Erik Castello Sections 3 and 4.2, and Viviana Gaballo Sections 2 and 4.1.

Note 2. However, some hairstyling manuals targeting non-native speakers (e.g., Canazza et al., 2020; Cascella, 2012) are available, which aim to develop readers' professional and linguistic competence, including lexico-phraseological and speech-act-oriented competence.

Note 3. *Bangs* (n) belongs to those polysemous words (e.g., *damage*/*damages*) whose meaning in the singular differs completely, a *bang* (n) being (1) a resounding blow, (2) a sudden loud noise (3) a sudden striking effect/a quick burst of energy/thrill, (4) an act of copulation / a sexual partner, (5) an exclamation mark. Retrieved 21.12.2023 from https://www.merriam-webster.com/dictionary/bangs.

Note 4. This is also comparable to Szemińska's (2014) notion of a terminographic system of interrelated dictionaries.

Appendix 1

List of the most frequent URLs that compose the *UK-Hairstyle* corpus:

https://freelancecorner.co.uk/

https://graziadaily.co.uk/beauty-hair/

https://hairboutique.co.uk/

https://hairdressers-near-me.co.uk/

https://www.hairbylisa.co.uk/

https://www.hairdo.co.uk/

https://www.hairdomobile.co.uk/

https://www.hairdosalon.co.uk/

https://www.hairdressers.co.uk/

https://www.hairdressersandstylists.co.uk/

https://www.hairmagazine.co.uk/

https://luxurylondon.co.uk/

https://markleeson.co.uk/

https://metro.co.uk/

https://www.express.co.uk/

https://www.glamourmagazine.co.uk/

https://www.gq-magazine.co.uk/

https://www.marieclaire.co.uk/

https://www.mirror.co.uk/

https://www.vogue.co.uk/

https://www.whowhatwear.co.uk/

https://www.wisteriaavenue.co.uk/

https://yougov.co.uk/

Appendix 2

List of the most frequent URLs that compose the *IT-Acconciatura* corpus:

https://archzine.it/bellezza/acconciature/

https://beauty.thewom.it/capelli/

https://dilei.it/bellezza/

https://donna.fanpage.it/

https://figarobarbiere.it/

https://immaginazioneelavoro.it/

https://k-max.it/blog/taglio-di-capelli-per-uomo-stempiato-istruzione-per-uso

https://lcparrucchieri.it/

https://moaparrucchieri.it/acconciatori/

https://viverepiusani.it/

https://www.acconciaturedavide.it/

https://www.adapformazione.it/

https://www.ademia.it/

https://www.ajparrucchieri.it/

https://www.amica.it/

https://www.beautydea.it/

https://www.beautydiary.it

https://www.beautytester.it/capelli/

https://www.bellezzasalute.it/capelli/

https://www.cfptrissino.it/percorsi-formativi/acconciatura/

https://www.donnaclick.it/

https://www.donnesulweb.it/bellezza/

https://www.hairling.it/

https://www.ilgentiluomo.it/

https://www.iodonna.it/

https://www.jeanlouisdavid.it/

https://www.marieclaire.it/

https://www.oggisposa.it/

https://www.palmerschool.it/

https://www.pettinatura.it/

https://www.theitaliantimes.it/

https://www.vogue.it/

https://www.vanityfair.it/

https://www.wikihow.it/

Free institutional Internet references and the language of Covid-19

Silvia Cacchiani[1]

[1] University of Modena and Reggio Emilia, Reggio Emilia, Italy

Correspondence: Silvia Cacchiani, Viale Allegri 19, Reggio Emilia, I-42121, Italia. E-mail: silvia.cacchiani@unimore.it

Abstract

The present paper is concerned with the inclusion in free English-language institutional internet reference works of terms used to discuss health, disease, treatments and medical breakthroughs in the context of COVID-19. The focus is on the glossaries that are available on the websites of the UK Parliament and UK Government, i.e. on credible and authoritative platforms that are in various ways intended to serve as seats for asymmetrical transfer and mediation of knowledge about their operations and services (Engberg and Luttermann 2014): *Coronavirus (Covid-19) Definitions* is the online interactive glossary published by the Office of National Statistics in 2021; the *COVID-19 glossary* is published by the Parliamentary Office of Science and Technology (POST). Cross-verification of wordlists and lexicographical treatement in selected dictionary entries is carried out vis-à-vis free and unlocked content from the Oxford Reference platform and the English Wiktionary. Integrating insights from the Function Theory of Lexicography (Bergenholtz & Tarp, 1995), and Wiegand's (1977 ff.) Actional-Semantic Theory of Dictionary Form, we are able to demonstrate that Coronavirus (Covid-19) Definitions and COVID-19 Glossary provide basic answers to the queries of lay users – which is in line with the government's social responsibility to pursue health promotion, prevention of ill health and health protection.

Keywords: Actional-Semantic Theory of Dictioanry Form, COVID-19 terms, English, free institutional internet glossary, Function Teory of Lexicography, lay-users

1. Introduction

When new concepts need to be named, there are usually three options: extending the meaning of existing words (and phrases) in neosemanticisms, coining a new word or phraseological unit based on existing rules, also complex words relying on grammatical and extra-grammatical morphology, and borrowing words and phraseological units (Durkin, 2009; ten Hacken & Panocová, 2020, adapted).

The World Health Organization declared the COVID-19 / SarS-COV-2 outbreak a pandemic on 11 March 2020 (WHO, 2020). They determined that it was "an established and ongoing health issue which no longer constitute[d] a public health emergency of international concern (PHEIC)" on 5 May 2023 (WHO, 2023).

The unprecedented circumstances of the COVID-19 pandemic have necessitated adopting terms from virology, public health, medicine and allied disciplines in general language usage, as well as coining and using new terms related to those aspects of life that have been impacted. This was part of the strenuous efforts put in medical journalism (Hettiarachchi & Noreen 2020, 38), official government communication and the press towards accomplishing related objectives – from increasing public awareness and disseminating knowledge about COVID-19 research, health and policies (Keshvari et al., 2018, p. 14), through promoting specific measures and counter fear, to providing round-the-clock news coverage of reports, statements and publications by national and international news agencies, health organizations, research centers, institutes and institutions. There was an urgent need for researchers and health professionals to share specialist information, and for professionals, organizations, local and national governments to reach out to the public at large.

With the web in power stage (Moor 2005), our knowledge search habits have increasingly changed. We have learnt to demand fast(er) access to more and preferably free information (Sunstein, 2008; Lorentzen & Theilgaard, 2012; Lew & de Schryver, 2014). This means sidelining active searches and long-established paying options like the authoritative Oxford Dictionaries, in print or online, which command trustworthiness and esteem based on a history of subsequent editions, taken-for-granted lexicographic and professional expertise and the publisher's unrivalled reputation. On the other hand, webpages on the front end of google search listings are automatically held to be credible and authoritative sources of knowledge (Sunstein, 2008; Lorentzen & Theilgaard, 2012). In the context of the Covid-19, browsing the internet for terms would sooner or later cross-refer users to free collective reference works like Wikipedia or Wiktionary. Another scenario could involve users browsing freely-available sources on non-institutional and institutional platforms or landing on such pages when located at the front end of Google search listings (e.g. the UK Government's official directory, Gov.uk, or the pages set up by the U.S. Food and Drugs Administration at Fda.gov).

This raises a number of issues. Online lexicographical works by amateur lexicographers may de provide thin if incorrect content (Lew & de Schryver, 2014), depart from lexicographic practice and be consistently inconsistent. Wiktionary may not be adequate for comprehension and non-professional online reference tools may boil down to poor glossaries that explain terms for content published on the particular website, or products and services on offer (Fuertes-Olivera, 2009).

The present paper is concerned with the inclusion in free institutional English-language internet reference works of words and phrases, also new coinages, used to discuss health, disease, treatments and medical breakthroughs in the context of COVID-19. Another goal concerns starting discussion about meaning descriptions. The focus is on the glossaries that are available on the websites of the UK Parliament and UK Government, i.e. on credible and authoritative platforms that are in various ways intended to serve as seats for asymmetrical transfer and mediation of knowledge about their operations and services (Engberg & Luttermann, 2014; Cacchiani, 2018a, 2018b): *Coronavirus (Covid-19) definitions* (CoD) is an online interactive glossary published by the Office of National Statistics (ONS) in 2021; the *COVID-19 glossary* (CoG) is published by the Parliamentary Office of Science and Technology (POST). Cross-verification of wordlists and lexicographical menaing paraphrases in selected dictionary entries is carried out based on free and unlocked content from Oxford Reference (FrOR) and the English Wiktionary (EN-Wik). The analysis is strictly qualitative and aims to lay the groundwork for evaluating the strengths and weaknesses of internet refences.

The paper is structured as follows. Section 2 provides some introductory remarks on the compilers, functions and target users of the reference works under consideration. Section 3 is devoted to methodology and framework of analysis. While we work in the tradition of the Function Theory of Lexicography (Bergenholtz & Tarp, 1995), we also adopt insights from Wiegand's (1977 ff.) Actional-Semantic Theory of Dictionary Form. Section 4 looks into the inclusion of COVID-19-related lemmas in CoG and CoD vis-à-vis FrOR and EN-Wik, in connection with their purported functions and relevant extra-lexicographical social situations. Section 4 offers some concluding remarks.

2. Free internet reference tools

In this section we look at the reference works under scrutiny through the lens of the Function Theory of Lexicography (Bergenholz & Tarp, 1995; Bergenholtz & Nielsen, 2006; Fuertes-Olivera, 2009; Agerbo, 2017). The typical extra-lexicographical social situation associated with Covid-19 consisted of lay-users surfing the net for utilitarian data that could fulfil their cognitively-oriented and possibly operative questions. In lexicographic terms, cognitively-oriented needs cover the acquisition of linguistic and semantico-encyclopaedic information ("What?") (Tarp, 2008; Bergenholtz & Bothma, 2011); when the user 's needs and questions are operative, the focus is on procedural information ("How to?") (Agerbo, 2017).

Coronavirus (Covid-19) definitions (CoD) is the UK government's interactive glossary (as per page name, *Coronavirus (COVID-19) Interactive Glossary*). (Note 1) It was published during the Covid-19 emergency by experts at the Office for National Statistics (ONS), or the UK's largest independent producer of official statistics and its recognized national statistical institute. Broadly, their mission is to collect, analyze and disseminate statistics which serves the public good, about the UK's economy, society and population. (Note 2) Particularly, CoD is the outcome of work carried out by ONS experts and associates working on the Health and Social Care core engagement theme, who take up recommendations from the Covid-19 Infection Surveillance Digital Advisory Board. (Note 3)

CoD is a multi-field, free institutional restricted (i.e. domain-focused) glossary that allows access to around 60 lemmas via a search box and drop-down menu. It inherits features that characterize the Gov.uk platform. In the words of the UK Government, Gov.uk is "[t]he best place to find government's services and information". (Note 4) Research into the digital written text and the visual representation of utility content on selected directories (Cacchiani, 2018a, 2018b) indicates that a significant portion of Gov.uk is designed for asymmetric communication of specialized knowledge from the UK government to lay citizens; the platform comes close to realizing mature information formats (Tognazzini, 2014) via recourse to usable (Nielsen, 1995 ff.) webpage layout and user interfaces especially intended to give citizens quick and easy help and support with utility content, or basic users' queries about knowledge and documentation that they might need. We therefore assume a tool that serves the knowledge-oriented needs of specific target users – semi-experts and lay-users – who are increasingly confronted with scientific and technical concepts and require assistance with comprehending information about the evolving Covid-19 scenario.

We do not assume that CoD's target users are experts or semi-experts in health and medicine, economics and social science. One reason for this appears to be that CoD is not connected via clickable buttons to expert content such as provided on Freedom of Information (FOI) (Note 7) – with pdf landing sites answering user requests of documentation, e.g. *Deaths from COVID-19 by vaccination status up from January to August 2023*, released on 20 September 2023. Additionally, CoD is not linked to the specialist datasets and publications on Coronavirus (COVID-19), (Note 5) on the social and economic impact of COVID-19.

A comparable resource appears to be the *COVID-19 glossary* (CoG). (Note 6) The resource was published on 13 January 2022 by the Parliamentary Office of Science and Technology (POST), whose mission is to source "reliable and impartial scientific research for Parliament." CoG comprises nearly 200 "definitions for the most commonly used scientific terms relating to COVID-19, as well as a list of organizations involved in public health, their acronyms and descriptions of their work."

Based on transmitter's descriptions, inclusion in UK institutional pages, as well as use of plain language in the digital written copy preceding the glossary and layout arrangements for information layering and scannability, it seems reasonable to suggest that the CoG caters for the comprehension needs of the general public. Observed usability features that aim to transfer, mediate (Engberg et al., 2018) and make utility content accessible include high Flesch Reading Ease scores (70 to 80 out of 100; as determined by the WebFX Readability Test Tool), as well as clauses with no more than 15 words (Loranger, 2017) and writing at the 6-8 Grade Level for the general audience and at up to Flesch-Kincaid Grade Level 12 for experts (Nielsen, 2015). Other important considerations are the use of in-page hyperlinks with items that do not overlap conceptually within each (sub-)category (Huei-Hsin & Chan, 2023) and recourse to interlocutive dialogic devices (Bres, 1985) such as imperatives and second-person pronouns for addressing users. Importantly, the glossary is also intended for assisting users with decoding and understanding POST contents: this is demonstrated by the *COVID-19* clickable button, which takes users to another location (Note 8) for summaries and full PDFs of research briefings, rapid responses and notes about aspects of health and social care related to Covid-19.

Oxford Reference – Answers with Authority (OR) (Note 9) is an entirely different tool. Libraries in the UK maintain subscriptions for public use to OR platform – which provides access to Oxford University Press's authoritative Dictionaries, Companions and Encyclopedias in 25 subject areas. Compiled by leading experts, their express function is to make specialized knowledge accessible to multiple target users. Generally, based on the outside matter (preface and promotional blurb), the titles that have been remediated for online can be said to serve as quick references for experts and professionals, as reference materials for those working in allied professions, as essential A-Zs for students in the subject field, as guides for professionals in other subject fields, and as answers to knowledge-oriented question that lay-users might have about various aspects of the subject and area of concern. We are exclusively interested in free and unlocked content (henceforth, Free Oxford Reference: FrOR) such as macrostructures, partial microstructures and full entries available in Overview. In particular, the emphasis lies in areas in the domain 'Medicine and Health', as well as in 'English Dictionaries'.

Wiktionary (Wik) is the prototype of a free multi-language internet dictionary that has been collectively compiled by people that do not belong to recognizable private or public organizations. All entries can be edited by users. Users and compilers are said to be "passionate about quality" (Note 10). Ideally, they follow the standard lexicographic protocol that is availbla in the outer matter. Their goal is to reach consensus within the community around accurate, neutral summaries of facts for prospective dictionary users (Fuertes-Olivera, 2009, p. 107). On the day it was last accessed (1 December 2023), the *English-language Wiktionary* (EN-Wik) includes a dictionary, "thesaurus, a rhyme guide, phrase books, language statistics and extensive appendices. Besides definitions, dictionary entries comprise information about etymologies, pronunciations, sample quotations, synonyms, antonyms and translations." (Note 11)

3. Methodology and framework of analysis

Our intention in this paper is to look into the macro- and microstructures of CoG and CoD in connection with their purported functions, compiler profiles, user needs and profiles, and the relevant extra-lexicographical social situations. Meaning representation within microstructures and, where present, mediostructures will be only addressed in passim. To this purpose, we work in the tradition of the Function Theory of Lexicography (Bergenholtz & Tarp, 1995). While we are fully aware of the many ways in which it has rejected, incorporated or modified other paradigms over time– for one, Wiegand's (1977 ff.) notion of *genuine purpose* of the dictionary (see Bergenholtz & Tarp, 2003) – we integrate basic tenets from the Function Theory with notions from Wiegand's (1977 ff.) Actional-Semantic Theory of Dictionary Form.

Addressing the needs of lay-users in free restricted institutional glossaries and in remediated utilitarian, commercial paper dictionaries for multiple target users has a bearing on wordlist, mediostructure and type and amount of information included in the entry. First, semasiological structures allow students, semi-experts and lay users to retrieve information, which they would not be able to look up via the underlying concept or related hyperonym. Regarding CoD and CoG, in terms of Search Engine Optimization, entries accessed via queries in the search box or the drop-down menu (as in CoD) might turn out offer short and perhaps thin content above-the-fold; entries in wordlists that extend beyond the current screen might prove to be even shorter and thinner. Turning to professional lexicographic resources like the books in OR, and, for that matter, FrOR, we expect entries in a semasiological macrostructure, with short but not deplorably thin content that users will visualize above the fold. Generally, for longer copy, modularity is a distinctive features of user-friendly dictionary entries that are the outcome of professional lexicography (e.g. OR dictionaries, though not CoD and CoG). That is, different senses

are listed in a nested structure. On the other hand, entry size is kept within reasonable limits thanks so specific features of the medio- and microstructure: at the mediostructural level, cross-referencing to related entries as a form of meaning description and further explanation of synonyms, hyperonyms, etc. or, at the microstructural level, condensation via standard dictionary conventions. Wiegand (1977 ff.) key terms are used to start discussion of lexicographical practice in dictionary entries (reference units; A [Wörterbüchartikel]) from CoD and CoG vis-à-vis FrOR and EN-Wik.

a) (Non-natural) *condensation* (*v.*: condensed [verdichtete]) reduction of form, e.g. via ellipsis, summary, substitution, shifting, abbreviation, etc.

b) *Lexicographical definition*: a text made up of *definiendum*, *definitor* (definition copula or absent relational expression), *definiens* (meaning paraphrase; BPA [Bedeutungsparaphraseangabe]).

c) Discrete *functional text segment* within the entry: (basic) reference unit – either *item* or *sentence* – that can be identified based on function (and position); functional additions are enlarging text segments that are not separable in a functional positional way. For instance, usage glosses, specifications of the reference domain in the definiens (domain labels, FGA [Angabe des Fachgebiets]), specifications of the referent (BezSPA [Angabe des Spezifizierung des Bezugsobjects]), or, expanding on that, of other elements within a specific frame (FrSPA [Angabe des Spezifizierung der Frame]). Elementary segments only have one segment and one function. Non-elementary segments can have homogeneous or heterogeneous segments.

d) *Lexicographical meaning description*: expository text that helps solve communication problems via answering predictable questions from prospective users (Wiegand, 1985: *user prerequisite principle*).

e) *Frame-based entry structure*: with answers about basic descriptors in the definiens, which denote *knowledge of categories* (F(K) [Kategorie (Wissen der Kategorie)]), *function* (K(F) [Kategorie (Wissen der Function)]), *forms and components* (K(FBT) [Kategorie (Wissen der Form und Bestandteile)]), *materials* (K(M) [Kategorie (Wissen des Materiales)], also causes).

f) *Hierarchical microstructure*: the structure of the reference unit, comprising information of *equivalents* (ÄquA [Angabe zur Äquivalenz]) and *comments* and *subcomments* (K [Kommentar], SK [Subkommentar] on *semantics* (SSK [Subcommentar zur Semantik] in segments *on meaning* (BA [Bedeutungsangabe]) such as *synonym* (SynA [Synonymeangabe]), *meaning paraphrases* (BPA), *examples* (KBeiA [Kompetenzbeispielangabe]), etc. Segments giving *polysemy*, as in nested structures, are specified by the acronyms PA1, PA2 [Polysemieanganbe 1, Polysemieangabe 2], etc.

g) *Integrate core*: Segments giving meaning, which play a crucial role in meaning description. It may be followed by front or back integrates or comments, e.g. on *etimology* (EtyA [Angabe zur Etymologie]). *Semantico-encyclopaedic* information may be present, e.g. in semantico-encyclopaedic comments (sem-enzyK [Semantisch-enzyclopädisch Kommentar]).

h) *Enlargement*: BPAs may be enlarged in many ways. For instance, with mediostructural *cross-references* (VerwA [Verweisangabe]) to additional *specialist information* (AFE [Angabe zur fachlichen Erklärung]).

Regarding CoD, it is important to note that in line with usability guidelines for content layering and scannability, separate wordlists are arranged around conceptual domains within what we may refer to as the 'COVID-19 ICM' (sensu Lakoff, 1987: ICM: Idealized Cognitive Model), thus approaching: terms used to describe the biology of the virus (Bio), terms used in understanding how covid spreads and how it can be contained (Spr-Cont); terms used in research about COVID-19 (Res), e.g. in subfields such as epidemiology and immunology, biomedicine and biochemistry; terms about statistics (Stats), which are an essential part of research in (public) medicine and health, epidemiology and immunology, biomedicine and biochemistry; terms used in drug development (Dev) and in Covid-19 treatment (Treat); terms used to discuss the immune response to Covid-19, immunizations and vaccines (Resp); names of national and international organizations involved in the Covid-19 response, and in public-health and medicine regulations, decision-making and scientific advice in response to Covid-19 (Org).

The analysis in Section 4 is organized around the domains Bio, Spr-Cont, Res, Stats, Dev, Treat, Resp and Org. Matching tables will be presented in turn, with separate columns for the wordlists of CoD and CoG, OR, FrOR, and OD – i.e. free, unlocked and restricted entries in OR's English Dictionaries – as well as EN-Wik. Based on manual examination of all entries, they detail information around the descriptors in (a) to (c). Additional observations on selected lemmas within each category are given as the analysis unfolds. The section concludes with some remarks on lexicographical practice in the microstructures.

a) √: Inclusion within the wordlist of terms in the particular subfield, also multi-word expressions (√*Acr*: Acronym; √*FF*: Full form). Alternatively, *BPA*: Inclusion in the meaning paraphrase; *PAn*: Inclusion in nested entries; *Bei*: Inclusion in explanatory example; *FE*: Occurrence in items giving further specialist information; *Ver*: Internal/External cross-reference, e.g. to the English Wikipedia (Wip).

b) *R-Lemma*: Inclusion of terms that are paradigmatically related (*P*) to CoD and CoG lemmas via synonymy, hyponymy, hyperonymic and hyponymic relations, meronymy or holonymy (Murphy, 2016), or that can be associated to CoD and CoG lemmas based on external relations in the ICM (*ICM*) such as action types (e.g.

purposive 'is used for'), process types (e.g. originatory 'originates/is converted into'), etc. (Ruiz de Mendoza Ibáñez & Pascual Aransaez, 1997-1998); *Fo-Lemma*: inclusion of lemma with similar form.

c) *Gen*: Inclusion of general vocabulary. For polysemy, *BezSP*: Lemma with changed referent; *FrSP*: Lemma with different frame features; *FGASP*: Lemma in a different subject field/reference domain; *Met*: Metonymic semantic shift.

4. Data analysis and discussion

Table 1 presents an overview of the terms used to describe the biology of the virus (Bio). General references like OD and EN-Wik provide ample coverage of *DNA* and paradigmatically related terms (*gene*, *genome*, *RNA* and *mRNA/messenger RNA*), co-hyponyms like *pandemic* and *epidemic*, and new Covid-related words, such as *COVID-19* and *SARS-CoV-2* in the fields of public health, medicine and epidemiology. *Lineage*, *carriage*, *fitness* or *reservoir* are generic terms – there are no polysemy items that cover reference in the subfields of genetics and biochemistry/biology.

OD resources at OR dictionaries can assist lay-users with the comprehension of words and terms that have become more common with Covid-19. In this connection, *Garner's Modern English Usage* (2022, 5 ed.) states that *pandemic* has become significantly more frequent than before the emergency, "ubiquitous in fact" with particular reference to the Covid-19 epidemic (GME: *pandemic*). Crucially, however, coverage of new terms such as *COVID-19* and *SARS-CoV-2* in the nomenclature for respiratory diseases and their causing agent is limited in OR to *A Dictionary of Nursing* (2021, 8ed.; DN: *COVID-19*; *SARS-CoV-2*) and absent from FrOR, while *long COVID* and *spike protein* have not been recorded yet. Also, OR resources in specialist (sub)fields do not cover metonymic shifts in general language use. Consider *Coronavirus*: as per the Baltimore classification of viruses and enveloped viruses, in science *coronavirus* denotes "[a] family of viruses that have a positive-strand RNA genome and are characterized by a viral envelope from which petal-shaped spikes protrude. The virus causing severe acute respiratory syndrome (SARS) in humans belongs to this family. Its genome contains 27,727 nucleotides." (*A Dictionary of Genetics*, 2014, 8. ed.; DGn: *coronavirus*). On the other hand, in the Oxford English Dictionary online, 2 ed. and Additions, sense 2 of the *Coronavirus* entry – originally an addition from 2003, during the SARS-1 outbreak – has been rewritten to include circumstantial information about the more recent Covid-19, as well as to account for the metonymic shift from infective agent to the disease in general language use, possibly an instance of determinologization (sensu Meyer & Mackintosch, 2000): "Any of the coronaviruses (genus *Betacoronavirus*) responsible for outbreaks of life-threatening respiratory disease in humans, esp. the major pandemic beginning in 2019 (see Covid-19 *n.*). Also as a mass noun: the disease caused by such a virus; *spec.* Covid-19' (OED: *coronavirus*, n. 2)." (For discussion of the language of COVID-19 in the Oxford English Dictionary, see Salazar & Wild, 2022.)

Table 1. Covid-19-related terms in Bio

	CoD	CoG	OR	OD	FrOR	EN-Wik
Antigenic drift		√	√ Antigenic drift; Fo/Syn-Antigenic shift		√ Antigenic drift; Fo/Syn-Antigenic shift	
Antigen		√	√	√	R-Antigen-antibody immune response	Verw-Wip-Antigen
Carriage		√	PAn-Carriage	Bez-Carriage	Bez-Carriage	Bez-Carriage
Coronavirus	√		√	√; R-Coronavirus	√	√
COVID-19	√	√	√	√		√
DNA		√	√	√	√	√
Endemic disease		√	√	R-Endemic	√	R-Endemic
Epidemic	√		√	√	√	Verw-Wip-Epidemic disease
Fitness		√	√	Gen-Fitness	√	Gen-Fitness
Gene		√	√	√	√	√
Genome		√	√	√	√	√

Lineage		√	√; R-Cell lineage mutans	R-Cell lineage; Bez-Lineage	√; Cell lineage mutants	
Long COVID		√		√		√
mRNA		√	√FF-Messenger RNA	√FF-Messenger RNA	√; √FF-Messenger RNA	√
Mutation		√	√	√	√	√
Omicron		√		√		
Outbreak	√		√	√; Bez-Outbreak	√	√
Pandemic	√		√	√	√	√
Replication		√	√	Gen-Replication	√	√
Reservoir		√	√ Reservoir; R-Reservoir of infection	Bez-Reservoir		√
Reverse zoonosis		√	R-Zoonosis			
RNA		√	√	√	√	√
Sars-CoV-2	√	√	√; R-SARS	√	R-SARS	√
Shedding		√				
Spike protein		√		√		√
Variant		√	√	√	√	√
Variant Of Concern (VOC)		√	FG-VOC			√
Variant Under Investigation (VUI)		√	FG-VUI			
Viral load	√		√	√		√
Virology	√	√	√	√	√	√
Virulence	√	√	√	√	√	√
Virus	√		√	√	√	√
Virus Like Particles (VLPs)		√				√
Zoonotic disease	√	√	√ Zoonotic disease; R-FO-Zoonosis; F-Zoonotic viruses; R-Zoonotic transmission	√	R-Zoonotic transmission; R-Zoonotic viruses	R-Zoonosis-VERW-Zoonotic

Confirmation of the findings on trends in the coverage of terms in Table 1 find support in Table 2, which reviews the terms related to the spread and containment (Spr-Cont) of Covid-19. An important point concerns terms that were borrowed between specialist domains (Durkin 2009, p. 164: borrowing within languages, between different specialist registers) and/or have become more frequent than before in relation to covid. CoD and CoG offer access to one-word and multi-word units, including highly technical terms such as *False positive* and *False negative*, *Flattening the curve*, *Patient zero*, *Booster* and *Super-spreader*, which have only recently become more common in expert discourse and general language use.

Turning to borrowing between different specialist registers, in sociology and geography *Social distance* denotes the 'perceived distance between social strata (different socio-economic, racial, or ethnic groups), usually measured by the amount of social contact between group' (A Dictionary of Geography, 2023, 6 ed.; DGeo: *social distance*); the associated action, *Social distancing*, is first attested in the OED (*social distancing*, n. 1) in 1957. In health care it denotes '[t]he action or practice of maintaining a certain physical distance from, or limiting physical contact with, another person or people (esp. family and friends), in order to reduce the spread of an infectious disease. The neosemanticism, first recorded in OED in 2004 (*social distancing*, n. 2), was previously used in connection with the SARS-1 outbreak of 2002-2004 and came to prominence during the COVID-19 pandemic. On the other hand, *Shielding* (CoD) is defined with reference to physics and spectrometry, while *S-gene* (CoG: S-gene drop-out,

where S-gene is the gene that gives the spike protein its crown-like shape) is one of two *Sex-genes* in OR, FrOR, OD and EN-Wik.

Table 2. Covid-19 related terms in Spr-Cont

	CoD	CoG	OR	OD	FrOR	EN-Wik
Accuracy test		√	√; R-Accuracy	Gen-Accuracy	√; R-Accuracy	Gen-Accuracy
Aerosol	√		√	√	√	
Airborne transmission	√		FE-Indirect transmission; FE-Contact		FE-Indirect transmission; FE-Contact	
Antibody test	√	√	√	√	√	
Antigen test		√	√	R-Antigen	R-Antigen	
Asymptomatic	√	√	√	√	√	√
Attack rate		√	√		√	
Booster			√	√	√	
Case fatality ratio		√	Fo-Case-fatality rate	Fo-Case-fatality rate	Fo-Case-fatality rate	Verw-Wip-Fo-Case-fatality rate
Contact tracing	√	√	√	√	√	√
COVID-Status Certification		√	R-Certificate of vaccination; R-International certificates of vaccinations		R-Certificate of vaccination; R-International certificates of vaccinations	
Cycle threshold	√		R-Threshold	Gen-Threshold	R-Threshold	√
Diagnostic test		√	√; √FE-False negative test		√; √FE-False negative test	
Doubling time		√	√		√	
Droplet transmission		√	R-Transmission; Bei-droplet transmission			
Epidemiologist		√	√	√	√	√
Epidemiology		√	√	√	√	Verw-Wip-Epidemiology
Excess mortality		√				
False negative	√	√	√	√	√	√
False positive	√	√	√	√	√	√
Flatten(ing) the curve	√	√	R-Epidemic curve; (in Medicine and Health), R-Gaussian curve; R-Bell-shaped curve; R-J-shaped curve	√	R-Epidemic curve; (in Medicine and Health), R-Gaussian curve; R-Bell-shaped curve; R-J-shaped curve	√
Fomite		√	F-Fomes	F-Fomes	F-Fomes	√

Term						
Growth rate		√	√	RE-Bei-Growth rate	√	
Immunity passport		√		√		
Incidence		√		√		√
Incidence		√		√		√
Incubation period	√	√	√	√	√	√
Infection fatality ratio		√	Fo-Fatality rate	R-Gen-Fatality	Fo-Fatality rate	R-Gen-Fatality
LAMP test or RT-lamp test (Reverse Transcription Loop-mediated isothermal AMPlification)		√	R-Loop mediated isothermal amplification; LAMP; R-Reverse transcription		R-Reverse transcription	
Lateral flow test		√		Lateral flow		√
Lockdown	√		√	√		√
Mass spectrometry		√	√	√	√	√
Mass testing		√	R-Testing	R-Testing	R-Testing	
Molecular test		√				
Morbidity		√	√	√	√	√
Mortality		√	√	√	√	√
Non-pharmaceutical interventions (NPIs)		√	R-Upstream intervention; R-Early intervention; R-Nursing intervention; R-Crisis intervention	R-Pharmaceutical intervention	R-Upstream intervention; Early intervention; R-Nursing intervention; R-Crisis intervention	
Nose and throat swabs	√		R-Swab	R-Swab	R-Swab	R-Swab
PCR (Polymerase Chain Reaction) test		√	√-PCR; R-Multiplex PCR; R-RT-PCR	√	R-Multiplex PCR; R-RT-PCR	R-Polymerase chain reaction
Personal Protective Equipment (PPE)	√		√; √Acr-PPR	√	√	√
Pooled testing		√	√-Fo-Pooled comparison test; Pooling		√-Fo-Pooling; R-Active Pool	
Positivity rate	√		R-Positivity			√
Prevalence	√	√	√	√	√	√
Primary case		√	√	√		
Quarantine	√		√	√	√	√

R (Reproduction Number)		√	√			
R nought	√		Fo-R0; Fo-R0t	Fo-R0	Fo-R;-Fo-R0t	Fo-R0
Rapid test		√	R-Speed test; R-Rapid epidemiological assessment		R-Speed test; R-Rapid epidemiological assessment	
Real-time Assessment of Community Transmission (REACT)	√		R-Rapid epidemiological assessment	R-Community transmission	R-Rapid epidemiological assessment	
S-gene drop-out		√	PA-FG-S genes; R-G-Protein	Fo-PA-Sex gene		PA-FG-S genes
Saliva test		√	√	√	√	√
Self-isolation	√		R-Isolation; R-Strict isolation	√ Self-isolation; Gen-Isolation		√
Self-sampling		√	Fo-Self-selected sample; Fo-Self-selection sample		Fo-self-selected sample	
Sensitivity		√	√; R-PA-Sensitization	Gen-Sensitivity	√; R-PA-Sensitization	√
Sequencing		√	√	√	√	√
Serial interval		√	√		√	√
Shielding	√		FG-Shielding	FG-Shielding	FG-Shielding	FG-Shielding
Specificity		√	Fo-Specific	√GEN-Specific	Fo-Specific	√
Super-spreader		√			Fo-Superspreader	√
Swab test and self-swabbing		√				
Symptomatic	√	√	√	Gen-Symptomatic	Fr-Symptomatic treatment	
Test and trace	√	√				√
Transmissibility		√	Fr-Transmissible disease	√	Fr-Transmissible disease	√
Transmission		√	√	Gen-Transmission	√	√
Quarantine	√	√	√	√	√	√
Vaccine interval		√	R-Interval	Gen-Interval	R-Interval	Gen-Interval
Vaccine passport		√	R-Certificates of vaccination; R-International certificate of vaccination in travel and tourism		R-Certificates of vaccination; R-International certificate of vaccination in travel and tourism	√
Ventilation	√	√	√	√	√	√

Table 3 presents the very few terms in statistics (Stats) that were not included in other modules. What should not go unnoticed is that multi-word units such as *adjusted odds ratio* (CoD) and *credible interval* (CoG) are not represented in other resources.

Table 3. Covid-19 related terms in Stats

	CoD	CoG	OR	OD	FrOR	EN-Wik
Absolute risk		√	√		√	
Adjusted odds ratio		√		R-Odds ratio	R-Adjustment; R-Odds ratio	R-Odds ratio
Credible interval	√		R-Interval	R-Interval	R-Interval	
Odds ratio	√	√	√	√	√	√

Table 4, comprising terms in drug development (Dev), Table 5, organized around terms in Covid-19 treatment (Treat), and Table 6, with terms used to discuss the immune response to Covid-19, immunizations and vaccines (Resp), all align with the observations above. OR, EN-Wik, as well as the combined wordlists of CoD and CoG, provide coverage of one-word and multi-word units that have entered the general vocabulary with their referents (e.g. *Antibiotics*, *Anti inflammatories* and *Side effects*, as in patient information leaflets; *Immune response* from medicine and epidemiology). A case in point is *Herd immunity*: OED's earliest illustrative example with a human referent is from 1927 (OED: *herd immunity*), in connection to one of the most notorious outbreaks of diphtheria in the 1920s. Other terms in CoD and CoG are reductions of longer units in OR, e.g. *Adverse event* (*Adverse health event*), *Antibody therapy* (*Antibody-directed drug therapy*); *Antiviral* (*Antiviral drug*). Yet another case are head-modifier constructs with head and modifier, often with some form type of reduction, which are given as separate lemmas in OR, e.g. *Active component* (*Biologically active* and *Component*). And, finally, lemmas in CoD and CoG can be part of meaning descriptions, specialist comments or technical examples, e.g. *Recombinant protein-based vaccine* in OR's *Recombinant protein*, or *Live-attenuated vaccine* in OR's *Immunization.*

Table 4. Covid-19 related terms in Dev

	CoD	CoG	OR	OD	FrOR	EN-Wik
Adverse event		√	Fo-Adverse health event		R-Adverse effect; R-Critical incidents	
Antibody therapy		√	Fo-Antibody-directed drug therapy		Fo-Antibody-directed drug therapy	
Antiviral		√	√; Fo-Antiviral drug	√	Fo-Antiviral drug	√
Antiviral drug resistance		√	R-Drug resistance	R-Drug resistance	R-Drug resistance	
Antiviral prophylaxis		√	√	√	√	Verw-Wip-R-Prophylaxis
Convalescent plasma		√	R-plasma	Gen-Convalescent; R-plasma	R-Plasma	Verw-Wip-Convalescent plasma
Good Manufacturing Practice (GMP)		√	√		√	
Monoclonal antibody		√	√	√	√	√
Pharmacovigilance		√		√		√
Side effects		√	√	√	√	√

Table 5. Covid-19 related terms in Treat

	CoD	CoG	OR	OD	FrOR	EN-Wik
Anti inflammatories		√	√	√	√	Verw-Wip-Anti-inflammatory
Extracorporeal Membrane Oxygenation (ECMO)		√	√		√	
Molnupiravir		√				
Ronapreve		√				
Sarilumab		√			√	
Sotrovimab		√				
Tocilizumab		√			√	
Ventilator		√	√	Gen-Ventilator	√	√

Table 6. Covid-19 related terms in Resp

	CoD	CoG	OR	OD	FrOR	EN-Wik
Active component		√	R-Component; R-Biologically active	Gen-Component	R-Component; R-Biologically active	
Active immune response		√	R-Immune response	R-Immune response	R-Immune response	Verw-Wip-Immune response
Adenovirus-based vaccine		√	R-Adenovirus	R-Adenovirus	R-Adenovirus	Verw-Wip-R-Adenovirus
Adjuvant		√	√; R-Adjuvant therapy	√	√; R-Adjuvant therapy	Verw-Wip-Adjuvant
Adjuvanted vaccine		√	R-Fo-Adjuvant		R-Fo-Adjuvant	
Antibiotic		√	√	√	√	√
Antibody	√	√	√	√	R-Monoclonal antibody	√
Attenuated vaccine		√	√			
B cells		√	√	√	√	Verw-Wip-B-cell
Booster	√		√; Fo-Booster dose	√	√; FO-Booster dose	
Cold chain		√	√	FG-Cold chain	√	Verw-Wip-Cold chain
Correlate of protection		√				
Cytokines		√	√	√	√	√
Disease modifying vaccine		√	R-Modifier		R-Modifier	R-Disease-modifying
DNA-based vaccines		√	Fo-DNA vaccine		Fo-DNA vaccine	
Dosing interval		√	R-Dose; FG-Dose fractionation	R-Dosage	R-Dose; FGA-Dose fractionation	R-Dosage
Eradication of disease		√	√	Gen-Eradication	√	Gen-Eradication
Fill-finish		√				
Herd immunity	√	√	√	√	√	√
Immune response		√	√	√	√	√

	CoD	CoG	OR	OD	FrOR	EN-Wik
Immunisation		√	√	√	√	√
Immunity	√	√	√	√	√	√
Inactivated vaccine		√	Bei-R-Killed vaccine	R-Vaccine	Bei-R-Killed vaccine	R-Vaccine
Inactivated virus		√	FE-Killed vaccine		FE-Killed vaccine	
Live-attenuated vaccine		√	FE-Immunization			
mRNA vaccine		√	√			
Passive immunity		√	√	√	√	√
Primary course of vaccination		√	√	R-Vaccination	R-Vaccination	
Priority groups		√				
Recombinant protein-based vaccine		√	FE-Recombinant protein		FE-Recombinant protein	
Scientific Pandemic Influence Group on Modelling (SPI-M)	√					
Self-amplifying RNA		√				
Stabilizer		√	√	Gen-Stabilizer	FG-Stabilizer	Gen-Stabilizer
Thrombocytopenia		√	√			√
Thrombosis		√	√	√		√
Trial batches		√				√
T cells		√	√	√	√	√
Vaccination		√	√	√	√	√
Vaccine candidate		√				
Vaccine coverage		√				
Vaccine take up		√				
Vectored vaccines		√	Vector dictionary of public health			
Virus Like Particles (VLPs)		√				√
Whole inactivated viral vaccine		√	Bei-Immunization			

Table 5 returns names of antiviral medicines in CoD and CoG that were granted marketing authorization for COVID-19 indication in 2021 and 2022, but were available on the market earlier, like *remdesivir*. Possibly as a consequence of editorial decisions, they are not covered in OR. In a slightly different manner, Table 7 (Org) demonstrates that OR and EN-Wik only record names of international bodies and organizations (*ECDC*, *EMA*, *FDA*, *WHO*), while CoG and to a minor extent CoD offer ample coverage of names for national groups (*SAGE*), companies (*IVQUIA*, *ZOE*), committees, institutes and organizations (*NICE*), agencies (*Public Health England*, *PHE*) and departments (*DHSC*) involved in the Covid-19 response, and in public-health and medicine regulations, decision-making and scientific advice in response to COVID-19.

Table 7. Covid-19 related terms in Org

	CoD	CoG	OR	OD	FrOR	EN-Wik
BEIS						
CDC						
CEPI		√				

Term						
CHM		√				
CMO		√				
COVAX		√				
CSA		√				
DHSC		√				
ECDC		√				√
EMA		√				√
FDA	√	√	√FF/Acr-Food and Drug Administration (FDA)	√	√FF/Acr-Food and Drug Administration (FDA)	√
GAVI, the Vaccine Alliance		√				
GCSA		√				
IQVIA	√					
JCB		√				
JCVI		√				
Lighthouse laboratories		√				
MHRA		√				
NERVETAG		√				
NICE		√				
Porton Down		√				
Public Health Agency (Northern Ireland)	√		R-Health agency; R-Governmental public health agency		R-Health agency; R-Governmental public health agency	
Public Health England (PHE)	√	√Acr-PHE	R-Health agency; R-Governmental public health agency		R-Health agency; R-Governmental public health agency	
Public Health Scotland	√		R-Health agency; R-Governmental public health agency		R-Health agency; R-Governmental public health agency	
SAGE						
SPI-B						
SPI-M						
UK HAS – UK Health Security Agency						
VTF						
World Health Organization (WHO); WHO	√	√	√Acr-WHO; √FF- World Health Organization	√	√Acr-WHO; √FF- World Health Organization	√
ZOE	√					

One final point that we would like to consider concerns definiens. Whereas we cannot provide a systematic description of lexicographical definitions and menaing paraphrases for lack of space, we will offer some preliminary observations first, and then briefly look at the lemmas *Coronavirus*, *Coronaviruses*, *COVID-19*, *Variant* and *Variant of Concern (VOC)*.

Generally, reference units in CoD and CoG are extremely short, and nested structures with semantico-encyclopaedic subcomments and multiple subsenses for technical specifications are not an option. One exception is (1), *Variant of concern (VOC)* (CoG), with identification of function in the meaning paraphrase and specialist additions for different variant types (*Alpha*, *Beta*, *Gamma*, Omicron), in turn coming with relator (*also known as*), synonym/equivalent item, and encyclopaedic information (*It was first detected in England in December 2020*) or mediostructural reference:

<p align="center">Variant of concern (VOC)   (1)</p>

Variant that, following a risk assessment by expert committees, is believed to have the potential for causing more severe disease, more deaths, increased transmissibility, resistance to treatments, or evading immunity conferred by vaccination or previous infection.

Examples of VOC currently monitored by UKHSA are [FE/Bei]:

> *Alpha* –also known as B.1.1.7 or VOC-20DEC-01 [Syn/Äqu/FE]. It was first detected in England in December 2020 [sem-enzy].
>
> *Beta* –also known as B.1.351 or VOC-20DEC-02 [Syn/Äqu/FE]. It was first detected in South Africa in December 2020 [sem-enzy].
>
> *Gamma* –also known as P.1 or VOC-21JAN-02 [Syn/Äqu/FE]. It was first detected in Japan in travellers from Brazil in January 2021 [sem-enzy].
>
> *Delta* – also known as B.1.617.2 or VOC-21APR-02 [Syn/Äqu/FE]. It was first detected in India in April 2021 [sem-enzy].
>
> *Omicron* – also known as B.1.1.529 or VOC-21NOV-01 [sem-enzy]. See the Rapid Response: COVID-19: The Omicron Variant [Ver].

(OdG: *Variant of concern (VoC)*
[Length: 128 words; Flesch-Kinkaid Grade Level: 6.3]

Recourse to condensation is exceptional in CoG (examples 1-3) and is only rare in CoG (examples 4-6). The very few mediostructural references are one-way: they exploit orthographic conventions for internal cross-referencing in CoG (1: *Covid-19*, *The Omicron Variant*), and use a combination of dialogic mediostructural additions and orthographic conventions in CoD (4: *See also antigenic drift*).

Most entries have an integrate core with definition by genus and differentia, specification of knowledge of category and function, and items or sentences that serve as specialist competence additions. They illustrate the case of objective exposition, which does not address users directly (Bres 1985; Section 3). Whereas readability scores in some entries align with usability guidelines (Nielsen 2015), as in examples (1) and (3), Latinate terminology can reduce readability in texts that strive for condensation, brevity and conciseness. as in (2).

<div align="center">

Coronaviruses   (2)

</div>

A family of viruses [K] that cause respiratory and gastrointestinal illnesses in people [F].
(CoG: *Coronaviruses*)
[Length: 13 words; Flesch-Kincaid Grade Level: 11.4]

<div align="center">

COVID-19   (3)

</div>

Coronavirus disease [K] first recognised in 2019 [FE/enzy]. The disease caused by SARS-CoV-2 [M, based in the ICM]. (CoG: *COVID-19*)
[Length: 13 words; Flesch-Kincaid Grade Level: 6]

<div align="center">

Variant   (4)

</div>

As a virus replicates, it can accumulate mutations. A version of the virus with these mutations is called a 'variant'. Emergence of variants is a natural phenomenon. Most mutations have very little impact on the virus's properties, others facilitate transmission or infection of other species. See also antigenic drift.
CoD: *Variant*
[Words: 183; Flesh-Kincaid Grade Level: 11.4]

<div align="center">

Coronavirus   (5)

</div>

Coronaviruses are a family of viruses [K] that cause disease in people and animals [F]. They can cause the common cold [F-Hypo] or more severe diseases [F-Hyper], such as COVID-19 [Hypo/Bei/FE].
(CoD: *Coronavirus*)
[Words: 27; Flesh-Kincaid Grade Level: 8]

<div align="center">

COVID-19   (6)

</div>

COVID-19 is the name used to refer to the disease [K] caused by the SARS CoV-2 virus [M], which is a type of coronavirus [K]. The Office for National Statistics (ONS) takes COVID-19 to mean presence of SARS-CoV-2 with or without symptoms [.
(CoD: *COVID-19*)
[Words: 44; Flesh Reading Ease Score: 9.1]

Turning to EN-Wik, the usual criticisms about this collective enterprise apply (Fuertes-Olivera, 2009). The lexicographical treatment accorded to each term is consistently inconsistent, despite the inclusion in the outside matter of guidelines for entry layout in the English side. For instance, *Coronavirus* cross-refers externally to the English Wikipedia (Table 1). Additionally, information in the definiens and specification of paradigmatically related meanings is often incorrect, as in the *COVID-19* entry in (7). Setting aside other issues, it is important to note that the domain label in sense 2 specifies that a term with precise reference in the disease nomenclature (*COVID-19*) has undergone a metonymic shift in *virology* (and not in general language use) to denote the causing virus (*Severe acute respiratory syndrome-related coronavirus 2*). Turning to paradigmatic meaning relations and external assocaitions, the article cross-refers to so-called '*Synomyms*' *SARS-CoV-2* and *SARSnCoV*, which are in fact mutual equivalents – in the sense that *SARSnCoV* has been later named *SARS-CoV-2* – but are not synonyms of *Covid-19*. Likewise, it is not clear how *Coordinate* terms relate to *Covid-19* based on paradigmatic relations or external associations.

<div align="center">

COVID-19 (7)

</div>

[…]

1. (*pathology*) A disease caused by a coronavirus discovered in 2019, in a zoonotic pandemic starting in Wuhan, Hubei, China.

2. (*virology*, *metonymically*) Severe acute respiratory syndrome-related coronavirus 2; the virus which causes the disease.

Synonyms

• (virus): SARS-CoV-2

• (virus): 2019-nCoV

Coordinate terms

(*disease*):

• MERS

• SARS

(*virus*):

• MERS-CoV

• SARS-CoV

(EN-Wik: *COVID-19*)

Altogether, lay users find basic answers to their questions on CoG. For semantico-encyclopaedic content and technical detail, they can look up FrOR content, which returns full and partial entries like (8).

<div align="center">

coronavirus (8)

</div>

any of a group of RNA animal viruses [K] consisting of enveloped particles 80-120 nm long, with helical nucleocapsids [FBT]. They contain the largest known viral RNA genomes (17-31 kb) and cause devastating epizootics (of respiratory or enteric disease) [FBT/M] in livestock and poultry [Bez]. Human [Bez] coronaviruses cause upper respiratory infections and severe acute syndrome (SARS) [Äqu] [F]. [sem-enzy/FE] See MAIN PROTEASE. [Verw]

(*Oxford Dictionary of Biochemistry and Molecular Biology*, 2006, 2 ed.; DBMB: *Coronavirus*)

[Free content; Full entry; Length: 60 words; Flesch Kincaid Grade Level: 12.2]

5. Conclusions

This chapter has offered a review of the coverage of terms related to Covid-19 in English language free institutional Internet glossaries that are available on the websites of the UK Government and Parliament – i.e. on credible and authoritative platforms that are in various ways intended to serve as seats for asymmetrical transfer and mediation of knowledge about their operations and services: *Coronavirus (Covid-19) definitions* (CoD) is the online interactive glossary published by the Office of National Statistics (ONS) in 2021; the *COVID-19 glossary* (CoG) is published by the Parliamentary Office of Science and Technology (POST). The written digital text in the glossary is based on research and statistics carried out by experts working for the government and parliament at the Office of National Statistics and the Parliamentary Office for Science and Technology.

Cross-verification of wordlists in CoD and CoG as well as lexicographical treatment in selected dictionary entries was carried out vis-à-vis free and unlocked content from the Oxford Reference (OR) platform and the English Wiktionary (EN-Wik), in connection with their purported functions and relevant extra-lexicographical social situations. To this purpose, we integrated somewhat liberally insights from the Function Theory of Lexicography (Bergenholtz & Tarp, 1995), and Wiegand's (1977 ff.) Actional-Semantic Theory of Dictionary Form.

The anlaysis demonstrates that the combined wordlists of CoD and CoG provide ample coverage not only of new terms such as *Covid-19* and recent terms that have not been recorded in OR yet. More specifically, CoD and CoG provide access to one-word and multi-word units that constitute reductions of longer strings in subject (sub)fields, e.g. *Adverse event* (*Adverse health event*), *Antibody therapy* (*Antibody-directed drug therapy*); *Antiviral* (*Antiviral drug*). Also included in the glossaries are head-modifier constructs with head and modifier, which are given as separate lemmas in OR, e.g. *Active component* (*Biologically active* and *Component*); lemmas that are part of general vocabulary (*Antibiotics*, *Anti-inflammatory drugs*); extant terms that have become to be used especially frequently (*Herd immunity*), and terms that were borrowed between specialist domains (*Social distancing*, *Shielding*). Additionally, we could observe ample coverage of names, from antiviral medicines eventually authorized for the treatment of Covid-19 (*Remdesivir*), to names for national groups (*SAGE*), companies (*IVQUIA*, *ZOE*), committees, institutes and organizations (*NICE*), agencies (*Public Health England*, *PHE*) and departments (*DHSC*) involved in the Covid-19 response, and in public-health and medicine regulations, decision-making and scientific advice in response to COVID-19.

Overall, Coronavirus (COVID-19) Definitions and the COVID-19 Glossary depart from professional lexicographical practice in a number of ways (e.g., in relation to non-natural condensation and mediostructural cross-referencing). However, we hope to have minimally demonstrated that they offer thin though not incorrect content. In fact, they provide basic answers to the potential questions of lay-users – which is in line with the government's social responsibility to promote health, prevent disease and protect health.

References

Agerbo, H. (2017). Monofunctional and polyfunctional information tools with an operative function. *Lexicograpica*, *33*, 361-390.

Bergenholtz, & Bothma, H. T. J. D. (2011). Needs-adapted presentation in e-information tools. *Lexicos*, *21*, 53-77.

Bergenholtz, H., & Nielsen, S. (2006). Subject-field components as integrated parts of LSP dictionaries. *Terminology*, *2*, 281-303.

Bergenhotz, H., & Tarp, S. (1995) *Manual of specialised lexicography: The preparation of specialised dictionaries*. Amsterdam: Benjamins.

Bres, J. (1985). *Dialogisme et polyphonie. Approches linguistiques*. Bruxelles: de Boeck-Duculot.

Cacchiani, S. (2018a). The voice of the law on Gov.uk and Justice.gouv.fr: Good value to citizens and institutions? In Engberg, J., Luttermann, K., Cacchiani, S., & Preite, C. (Eds.), *Popularization and Knowledge mediation in the law. / Popularisierung und Wissensvermittlung im Recht* (pp. 117-148), Munster: LIT Verlag.

Cacchiani, S. (2018b). Webpage usability and utility content: Citizens' rights and the law on Gov.uk. *Iperstoria*, *12*(Fall-Winter), 192-205.

DBMB: Cammack, R., Atwood, T., Campbell, P., Parish, H., Smith, A., Vella, F., & Stirling, J. (2006). *Oxford Dictionary of Biochemistry*, 2 ed. Oxford: Oxford University Press. https://www.oxfordreference.com/display/10.1093/acref/9780198529170.001.0001/acref-9780198529170 (1 December 2023).

DGeo: Mayhew, S. (2023). *A Dictionary of geography*, 6 ed. Oxford: Oxford University Press. https://www.oxfordreference.com/display/10.1093/acref/9780192896391.001.0001/acref-9780192896391 (l December 2023).

DGn: King, R. C, Mulligan, P. K., & Stansfield, W. D. (2014), *A Dictionary of Genetics*, 8 ed. Oxford: Oxford University Press. https://www.oxfordreference.com/display/10.1093/acref/9780199766444.001.0001/acref-9780199766444 (1 December 2023).

DN: Martin, E., & MrFerran, T. 2021. *A dictionary of Nursing*, 8 ed. Oxford: Oxford University Press. https://www.oxfordreference.com/display/10.1093/acref/9780198864646.001.0001/acref-9780198864646 (1 December 2023).

Durkin, P. (2009). *The Oxfordg guide to etymology.* Oxford: Oxford University Press.

Engberg, J., Cacchiani, S., Luttermann, K., & Preite, C. (2018). Studying popularization in legal communication: Introduction and overview. In Engberg, J., Luttermann, K., Cacchiani, S., & Preite, C. (Eds.), *Popularization and Knowledge mediation in the law. / Popularisierung und Wissensvermittlung im Recht* (pp. 4-25), Munster: LIT Verlag.

GME: Garner, B. A. (2022). *Garner's modern English usage*, 5 ed. Oxford: Oxford University Press. https://www.oxfordreference.com/display/10.1093/acref/9780197599020.001.0001/acref-9780197599020 (1 December 2023).

Hettiarachchi, D., & Noordenm, N. (2020). The impact and challenges of medical journalism: A review. *University of Colombo Review*, *1*(1), 37-46.

Huei-Hsin, W., & Chang, M. (2023, November 17). 5 formatting techniques for long-form content. https://www.nngroup.com/entries/formatting-long-form-content/ (1 October 2023).

Keshvari, M., Yamani, N., Adibi, P., & Shahnazi, H. (2018). Health journalism: Health reporting status and challenges. *Iran Journal of Nursing and Midwifery Research*, *23*(1), 14-17.

Lakoff, G. (1987). *Women, fire and dangerous things. What categories reveal about the mind*. Chicago, IL: The University of Chicago Press.

Lew, R., & de Schryver, G.-M. (2014). Dictionary users in the digital revolution, *International Journal of Lexicography*, *27*(4), 341-359.

Loranger, H. (2017, October 8). Plain language is for everyone, even experts. https://www.nngroup.com/entries/plain-language-experts/ (1 October 2023).

Lorentzen, H., & Theilgaard, L. (2012). Online dictionaries – How do users find them and what do they do once they have?. In Fjeld, R. V., & Torjusen, J. M. (Eds.), *Proceedings of the 15th EURALEX International Congress* (pp. 654-660). Oslo: Department of Linguistics and Scandinavian Studies, University of Oslo.

Meyer, I., & Mackintosh, K. (2000). When terms move into our everyday lives: An overview of de-terminologization. *Terminology*, *6*: 111-138.

Moor, J. H. (2005). Why we need better ethics for emerging technologies? *Ethics and Information Technology*, *7*(3): 111-119.

Murphy, L. A. (2016). Meaning relations in dictionaries: Hyponymy, meronymy, synonymy, antonymy, and contrast. In Durkin, P. (Ed.), *The Oxford handbook of lexicography* (pp. 439-456). Oxford: Oxford University Press. 439-456.

Nielsen, J. (1995, November 15). Ten usability heuristics for user interface design. https://www.nnggroup.com/entries/10-usability-heuristics (1 October 2023).

Nielsen, J. (2015, November 15). Legibility, readability, and comprehension: Making users read your words. https://www.nngroup.com/entries/legibility-readability-comprehension/ (1 October 2023).

OED: *Oxford English Dictionary*, 2 ed. and Additions. Oxford: Oxford University Press. https://www.oed.com (1 December 2023).

OR: *Oxford Reference*. https://www.oxfordreference.com (1 December 2023).

ORDL: Law, J. (2022). *A dictionary of law*, 10 ed. Oxford: Oxford University Press. https://www.oxfordreference.com/display/10.1093/acref/9780192897497.001.0001/acref-9780192897497 (1 December 2023).

Ruiz de Mendoza Ibáñez, F. J., & Pascual Aransaez, C. (1997-1998). Conceptual schemas as propositional idealized cognitive models: In the search of a unified framework for the analysis of knowledge organization. *C.I.F.*, *23-24*: 257-270.

Salazar, D. & Wild, K. (2022). The Oxford English dictionary and the language of Covid-19. In Klosa-Kückelhaus, A., & Kernerman, I. (Eds.), *Lexicography of Coronavirus-related Neologisms* (pp. 11-26). Berlin: De Gruyter.

Sunstein, C. R. (2008), Democracy and the Internet. In van der Hoven, J., & Weckert, J. (Eds.), *Information technology and moral philosophy* (pp. 95-110). Cambridge: Cambridge University Press.

Tarp, S. (2008). *Lexicography in the borderland between knowledge and non-knowledge. General lexicographical Theory with particular focus on learner's lexicography*. Tübingen: Max Niemeyer.

ten Hacken, P., & Panocová, R. (2020). Word formation, borrowing and their interaction." In ten Hacken, P., & Panocová, R. (Eds.), *The interaction of borrowing and word formation* (pp. 3-14). Edinburgh: Edinburgh University Press.

Tognazzini, B. (2014, March 5). First principles in interaction design (revised and expanded). https://www.ngroup.com/entries/first-principles-in-interaction-design/ (1 October 2023).

WebFX Readabiliy Test Tool: https://www.webfx.com/tools/readable/(1 December 2023).

Wiegand, H. E. (1977). Einige grundlegende semantisch-pragmatische Aspekte von Worterbücheinträgen. Ein Beitrag zur praktischen Lexikologie. *Kopenagener Beiträgen zur Germanistischen Linguistik*, *12*: 59-149.

Wiegand, H. E. (1992). Elements of a theory towards a so-called lexicographic definition. *Lexicographica*, *8*: 175-289.

Wiegand, H. E. (2015). Funktionale Angabezusätze und hybride textuelle Strukturen in Wörterbuchartikeln. Korrigierende Ergänzungen, weiterführende Vertiefungen, zusammenfassende Übersichten und mehr. *Lexicographica*, *31*(1): 249-330.

Wikipedia: https://en.wikipedia.org/wiki/Main_Page (1 December 2023).

World Health Organization. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020*. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 (1 October 2023).

World Health Organization. (2023). *Statement on the Fifteenth Meeting of the IHR (2005) Emergency Committee on the COVID-19 pandemic*. https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic (1 October 2023).

Notes

Note 1. https://www.ons.gov.uk/visualisations/dvc988/line/index.html (1 December 2023).

Note 2. https://www.ons.gov.uk/aboutus/whatwedo (1 December 2023).

Note 3. https://www.ons.gov.uk/aboutus/transparencyandgovernance/covid19infectionsurveillancedigitalcisdadvisoryboard (1 December 2023).

Note 4. https://www.gov.uk (1 December 2023).

Note 5. https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi (1 December 2023).

Note 6. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases (1 December 2023).

Note 7. https://post.parliament.uk/covid-19-glossary/ (1 December 2023).

Note 8. https://post.parliament.uk/tag/covid-19/ (1 December 2023).

Note 9. https://www.oxfordreference.com (1 December 2023).

Note 10. http://en.wiktionary.org/wiki/wiktionary:Welcome%2Cnewcomer (1 December 2023).

Note 11. https://en.wiktionary.org/wiki/Wiktionary:Main_Page (1 December 2023).

Idiomatic Toponymic Phrasemes in Bajan: Evidence from Amateur Paremiography

Cristiano Furiassi[1]

[1] Department of Foreign Languages, Literatures and Modern Cultures, University of Turin, Turin, Italy

Correspondence: Cristiano Furiassi, Department of Foreign Languages, Literatures and Modern Cultures, University of Turin, Via Sant'Ottavio 18, 10124 Turin, Italy. E-mail: cristiano.furiassi@unito.it

Abstract

This study begins with the quantification of the paremiological inventory of Bajan, the English-based creole spoken in Barbados, by examining all the printed collections of proverbs, both amateur and professional, available to date. Consequently, by relying specifically on the amateur paremiographic work on Bajan compiled in 1987 by G. Addinton Forde, namely *De Mortar-pestle: A Collection of Barbadian Proverbs*, this article examines not just the proverbs but also the phrasemes which employ territorial place names found therein. In greater detail, this investigation is intended to show how the idiomatic use of such toponymic phrasemes yields cultural representations characterizing the unique worldview of Barbadians both within and outside the Anglophone Caribbean. Despite the fact that the colonial history of the island and the recollection of the slavery era are still undeniably evident in the toponymic Bajanisms considered, whose authentically local components are almost entirely drawn from the names of former plantations, the intrinsic Barbadianness that stands at their core is foregrounded.

Keywords: Bajan, idioms, paremiography, paremiology, phrasemes, proverbs, toponyms

1. Introduction

It is axiomatic that proverbs are widespread in virtually all languages of the world (see Mieder, 2012, p. 108). However, there are languages in which they seem to be particularly prominent. As shown by recent research, a case in point is that of Caribbean Englishes and Caribbean English-based creoles (see Allsopp R., 2000/2006; 2004; Allsopp J. & Furiassi, 2020), respectively including Barbadian English and Bajan (see Furiassi, 2022; 2023), all characterized by a fairly large number of proverbs.

The present study must be placed within the wider framework of amateur paremiography, that is the compilation of "proverb collections" and "proverb dictionaries" (Kispál, 2015, pp. 229-230) by lay or vernacular lexicographers, a theme which has received academic consideration only recently. Accordingly, among the latest studies on the topic, Lambert (2020), Furiassi (2022; in press), Finegan (2020), J. Allsopp (2009) and Winer (2006) deserve to be mentioned.

Moreover, this piece of research delves into the highly restricted semantic field of geographical names, i.e. toponyms, as part of idiomatic phraseological units, including proverbs, a subject on which scholarly works appear sporadic. As for the investigation of toponyms and how they combine with other lexical elements to generate prefabricated pieces of language with an idiomatic meaning, the essays by Dueck (2004), Szerszunowicz (2009) and Bredis and Lomakina (2019) are worthy of recognition.

After offering a contextualization of Bajan, the English-based creole of Barbados, and providing an operational definition of idiomatic toponymic phrasemes, this article takes two parallel paths, respectively a quantitative and a qualitative one. While reviewing the Bajan paremiographic sources, both "amateur" and "professional" (Lambert, 2020, pp. 411-412), available to date in print, the focus shifts to the manual retrieval and the attendant counting of entries, so as to measure their numerical contribution. Subsequently, following the extraction of Bajan-only proverbs and other phrase-like units from the examined collections, this article centers on the phrasemes containing territory-specific toponyms precisely drawn from the primary source of this investigation, namely Forde's (1987) *De Mortar-pestle: A Collection of Barbadian Proverbs*, which, regardless of its amateur essence, represents the most complete paremiographic inventory addressing the salient features of Bajan proverbs. The typically Bajan toponymic phrasemes found therein are isolated and commented on with a view to explaining their idiomatic use and emphasizing their cultural relevance (see Dobrovol'skij & Piirainen, 2022, pp. 272-275).

2. Hints on Bajan

Several studies on Bajan have been carried out over the years (see Alleyne, 1971, p. 181; Rickford, 1992, p. 195; Fenigsen, 1999, pp. 65-66; 2003, p. 461; 2007, p. 233; 2011, pp. 111-112; Van Herk, 2003, pp. 241-243; Blake, 2004, p. 501; Schneider, 2007, p. 224; Haynes-Knight, Evans, & Winters-Evans, 2015, pp. 314-315; Stuka, 2023,

pp. 94-95) and a summary of both Bajan and Barbadian English has been recently provided by Furiassi (2022, pp. 91-92; 2023, pp. 49-51; in press). For the scope of the present analysis, suffice it to say that *Bajan*, the "clipping or shortening" of the adjective *Barbadian* (*Oxford English Dictionary Online*), identifies the English-based creole still spoken as a native language by the majority of residents in the Caribbean island of Barbados.

From a taxonomic perspective, "Barbadian Creole English" (*Ethnologue*) has been classified as a "mesolectal" (Rickford & Blake, 1990, p. 258; Gibson, 1996, p. 41) or "intermediate" (Winford, 2000, p. 215; 2003, p. 313) creole because of its remarkable lexical similarities with English, its superstratum language. Indeed, this new language creation draws its vocabulary from diatopic varieties of the Modern English of British colonizers and the 17th-century Irish English—and, to a lesser extent, the 18th-century Scottish English—of indentured servants who were deported to Barbados. In addition, inputs from the diastratic register of the lower classes and naval vernacular are not to be dismissed.

However, and most importantly, Bajan is also the offspring of various Niger-Congo languages—especially those presently in use in the contemporary West African states of Ghana and Nigeria—spoken by the enslaved peoples who were forced to work in the many plantations spread over the island. As Bajan and Barbadian English, the variety of English which is the national language of Barbados, work along a continuum, diglossia is evident in most cases: the former is usually encountered in colloquial registers and tends to be limited to orality; the latter, appearing more often in writing, is allocated to formal settings.

3. Terminological Issues: An Operational Definition of Idiomatic Toponymic Phrasemes

Traditionally, paremiography may be subdivided into "proverb collections" and "proverb dictionaries". On the one hand, "proverb collections […] interpret proverbs in a broader sense" by also including "proverbial comparisons", "wellerisms" and "even idioms" (Kispál 2015: 229) (Note 1). On the other hand, "proverb dictionaries […] interpret proverbs in a narrow sense and so they codify only proverbs that are generally sentential statements" (Kispál, 2015, pp. 229-230).

Since all the publications reviewed (see Section 4) include the term "proverb" in their titles, they are expected to record only proverbs in their respective wordlists. However, it is apparent that they also incorporate other types of phraseological material. This statement is certainly devoid of novelty since, according to Kispál (2015, p. 240), regardless of the theoretical differences mentioned, both proverb collections and proverb dictionaries are known to cover various kinds of phraseological items in addition to proverbs proper.

Even though, from a formal viewpoint, the sentence-like or utterance-like dimension of proverbs is emphasized in authoritative definitions (see Mieder, 1996, pp. 597; Klein & Lamiroy, 2016, p. 17-19), as noted by Norrick (2015, p. 14), "[w]ith culturally determined items like proverbs […] it is necessary to recognize the fuzziness of the category and the scalar application of features" (Note 2). Due to the mixed nature of the paremiographic sources under scrutiny, the scope of this analysis must be extended to the superordinate category of phrasemes (see Section 6).

In fact, proverbs themselves are just one of the several constituents—possibly the most central ones (see Piirainen, 2008, p. 214) —of the wide-ranging category of phraseology. Therefore, drawing on various studies on the matter, such as Cowie, Mackin and McCaig (1993, p. x), Mel'čuk (1995, p. 217), Cowie (1998, p. 9), Nuccorini (2006, pp. 37-38; 2016, pp. 60-61), Gries (2008, p. 6) and Pulcini, Furiassi and Rodríguez González (2012, p. 13), this analysis is based on the all-inclusive concept of "phraseme" (Mel'čuk, 1995, p. 179; 2012, p. 32), that is any multi-word expression formed by at least two syntactically linked lexemes which operates as one semantic unit in a clause or sentence. All in all, the definition of phraseme is in turn rather broad and, likewise somewhat characterized by fuzzy borders as it encompasses a continuum of different chains of readymade linguistic material, such as collocations, binomials, phrases, similes, idioms, catch phrases, routine formulas, sayings and proverbs, all subject to rather stable pragmatic constraints.

In particular, within the overarching class of phrasemes, those which rely on toponyms are analyzed from a qualitative perspective in the second part of this article (see Section 6). Before delving deeper into Bajan-specific phrasemes based on local place names, it is necessary to clarify what idiomatic toponymic phrasemes are by providing an operational definition of this concept. The starting point is represented by the term "toponym", or "place name", described by McArthur, Lam-McArthur and Fontaine (2018) as "[t]he proper name of a locality, either natural (as of bodies of water, mountains, plains, and valleys) or social (as of cities, counties, provinces, nations, and states)". Consequently, it is of paramount importance to consider the notion of "toponymic idioms" (Szerszunowicz, 2009, p. 172), namely semantically loaded idiomatic expressions containing geographical denominations which function as carriers of local identity. Thus, by drawing inspiration from the explanations offered above, idiomatic toponymic phrasemes may be defined as multi-word units constituted by place names equipped with culturally relevant connotational meanings alongside their denotational ones.

4. Barbadian Paremiography

A thorough description of Barbadian English and Bajan amateur lexicography (and lexicographers) has already been proposed by Furiassi (in press). As for Bajan paremiography, that is the compilation of proverb dictionaries pertaining to the English-based creole spoken in Barbados, the three amateur products printed to date are the following: Blackman's (1985) *Bajan Proverbs*, Forde's (1987) *De Mortar-pestle: A Collection of Barbadian Proverbs* and Ibekwe's (1998) *Wit & Wisdom of Africa: Proverbs from Africa & the Caribbean*. However, this survey would not be complete without also mentioning the two professional paremiographic products published thus far, namely Richard Allsopp's (2004) *A Book of Afric Caribbean Proverbs* (*BACP*) and Bogle's (2020) *The Transatlantic Culture Trade: Caribbean Creole Proverbs from Africa, Europe, and the Caribbean*. It is worth underscoring once more that, even though the collections analyzed below allegedly include only proverbs, all instances of phraseology encountered therein are considered while calculating the total number of entries.

After a collection period which spanned over almost thirty years, Margot Blackman's *Bajan Proverbs* came out in 1982 and was soon followed by a second edition in 1985, along with four reprints in 1987, 1989, 1992 and 1995. Blackman's paremiological glossary consists of eighteen pages and includes a bare list of 382 entries presented in alphabetical order. Unfortunately, the fact that the proverbs she catalogues are unaccompanied by any other piece of information makes it particularly difficult for the reader to fully understand the meaning of at least some of them. As a case in point, *Yuh t'ink I is a Moojun cow?* (Blackman, 1985, p. 18) is presumably opaque to native or proficient speakers of English. Luckily, the *DCEU* records the typically Bajan noun *moojin*—an alternative spelling variant of *moojun* or *moojink*—and defines it as "[a] fool; a stupid person; a worthless idiot". Here, the premodifier *moojun* is combined with the noun *cow*, which, interpreted as derogatory, refers to "a woman who is stupid or annoying" (*Merriam-Webster's Collegiate Dictionary*). All considered, uttering this idiom is tantamount to saying *Do you think I'm an idiot?* or *Do you think I'm stupid?*.

Almost four decades ago, in 1987, G. Addinton Forde compiled *De Mortar-pestle: A Collection of Barbadian Proverbs*, the most detailed repository of Bajan proverbs ever published. The amateur nature of Forde's achievement is confirmed by Marshall (1987, p. v), who, in the foreword, writes that "Addinton Forde […] worked on this collection as a silent toiler without the official encouragement or promise of assistance in publishing". Notwithstanding the wealth of microstructural information included therein, Forde's glossary, comprising a total of 295 main entries, is admittedly based—and elaborates—on a selection of headwords from Blackman's (1985) *Bajan Proverbs* (see Forde, 1987, p. viii). Forde's is a thematically organized proverb collection divided into three macro-categories: proverbs, similes and sayings (Note 3). Within each topic to which proverbs are assigned, they are arranged in alphabetical order. As opposed to Blackman's (1985) lack of proper lexicographic details, Forde's incorporates definitions and, in some cases, explanatory notes, often supplied with drawings and maps aimed at expanding upon the meaning or the etymology of said proverbs.

Wit & Wisdom of Africa: Proverbs from Africa & the Caribbean, a 4,000-entry collection assembled by Patrick Ibekwe during a ten-year period and published in 1998, consists of a mere list of proverbs gathered from different languages belonging to both Africa and the Caribbean. In 2009, select headwords from this publication were incorporated in *The Little Book of African Wisdom* by the same author. Within Ibekwe's collection, only 51 entries, marked as "BAJAN", are of interest for the present analysis.

The only major professional attempt at systematizing proverbs of the Anglophone Caribbean—including Barbadian English and Bajan—is *A Book of Afric Caribbean Proverbs* (*BACP*), published by Richard Allsopp in 2004. Even though the *BACP* includes as many as 1,313 proverbs, only 182 of them, tagged as "Bdos", are said to be current in Barbados. More precisely, while 105 of these are also used in other Caribbean English varieties or English-based creoles, as few as 77 pertain to Barbadian English and/or Bajan.

Authored by Desrine Bogle, *The Transatlantic Culture Trade: Caribbean Creole Proverbs from Africa, Europe, and the Caribbean* was published in 2020. Despite being the only other printed work professionally compiled, Bogle's glossary, which contemplates both Anglophone and Francophone Caribbean proverbs, holds almost no interest for Barbadian paremiography. As a matter of fact, only 5 of the 222 proverbs included in this collection are tagged as "Barbados" and thus belong to Bajan, namely, in alphabetical order, *De las' calf kill de cow*, *Last pickney kill mumma*, *Rat mout does sell he head*, *Two head-bulls can't reign in a flock* and *You can run but you can't hide*—the last one found in Antigua as well.

After pondering over the above remarks, it was decided to consider Forde's collection as the source of data for the qualitative portion of this investigation. Although it might not be the most numerous and despite being substantially amateur in nature, it is in fact the one that appears to be the closest to professional standards by offering a rather comprehensive treatment of entries. Additionally, it is particularly relevant as it incorporates the largest number of Bajan toponymic phrasemes, among which proverbs, characterized by idiomatic nuances. Indeed, no idiomatic phraseme inspired by Barbadian place names is found in Ibekwe (1998), Bogle (2020) or the *BACP*. As for Blackman's (1985) glossary, only the following three entries—all discussed below (see Section 6)—show the presence of toponyms: *All yuh labour gone in Maxwell pond* (Blackman, 1985, p. 1), *De Lazaretto*

dog die at de lighthouse (Blackman, 1985, p. 2) and *He'll promise you Drax Hall and Kendal* (Blackman, 1985, p. 5).

5. Data Source and Methodology

Although, as the title says, *De Mortar-pestle: A Collection of Barbadian Proverbs* would lead readers to think that it only contains proverbs from Barbados, Forde (1987, p. x) himself explains that "the term 'Barbadian proverbs' refers more to the usage than the origin". In fact, he further specifies that, at times, the proverbs listed in his collection are spread across other Caribbean territories in addition to Barbados—the same caveat being present in Blackman's (1985) glossary (see Collymore, 1966, p. 158; Brooks, 1985, p. viii).

Therefore, two main issues must be addressed if a reliable quantitative analysis of Forde's work is to be carried out. First, in addition to the 295 items listed as main entries within the collection (see Section 4), lexical variants must be added. Next, of the overall number of entries detected, only those belonging exclusively to Bajan should be considered.

As for the former, lexical variants are to be intended as proverbs which, regardless of their formal dissimilarities due to word choice, in fact convey the same meaning. Identifying synonymic proverbs in Forde's (1987) glossary turned out to be a complex task as they are either not signaled or, when they are, they are not identified by means of unvocal labels. More accurately, 20 proverb entries are preceded by "Also", 6 are introduced by "or" and 1 remains untagged, hence totaling 27 lexical variants, such as, for instance, *Wha' evah in de ole goat in de kiddie* and *De berry don' fall far from de tree*—both indicating that "[c]hildren inherit the traits of their parents" (Forde, 1987, p. 8). By adding these 27 variants to the 295 proverbs originally listed as main entries, the updated overall figure is 322.

As far as the latter aspect is concerned, the following procedures were implemented to select only Bajan-specific proverbs—and, more generally, phrasemes, hence excluding those which are allegedly widespread across other English-based Caribbean creoles. First, as many as 83 entries marked as "(Also) used in […]" were excluded from the overall count. For instance, *Fisherman never say dat 'e fish stink*, meaning that "[p]eople never give bad reports about themselves" (Forde, 1987, p. 12), is also current in Belize and Jamaica. Moreover, 2 entries which, below the definition, mention at least another Caribbean territory where the same proverb is said to be used, were likewise discarded. For example, *A cat can look at de queen*, referring to the fact that "[h]owever humble he is, every person has certain inalienable rights" (Forde, 1987, p. 1), is current in Grenada in its identical form. Therefore, if Forde's findings are deemed sufficiently reliable, by disregarding these 85 instances, the final figure obtained, that is 237, indicates the exact number of proverbs pertaining exclusively to Bajan.

6. Bajan-specific Toponymic Phrasemes

Amidst this rather extended list of Bajan phrasemes, it was decided to focus and comment on the (only) seven based on Barbadian place names—among which three synonymic alternatives were found: *All me labour gone in Maxwell Pond* (and its variant *All o' me labour gone in Lowther's Pond*), *De Lazaretto dog dead at Ragged Point Lighthouse*, *Dem does work at Idle Hall*, *Dis place hot as Mapp's millyard* (and its variant *Dis place hot as Bayley's millyard*), *Ent she hard! She just like Bulkeley iron man*, *He gine promise you Drax Hall and Kendal*, and *You don't tek down Drax Hall to fix Kendal* (and its variant *You can' tek down Drax Hall to put up Kendal*).

These toponymic phrasemes are highly conspicuous as not only are they directly connected with Barbadian realia but they also display a meaning which goes beyond the literal (Note 4). Needless to say, the toponymic phrasemes taken into account cannot be encountered anywhere else: they are found in neither the *Oxford English Dictionary Online* nor the *Merriam-Webster Unabridged Online*, in all likelihood the most trustworthy contemporary dictionaries encompassing the two best-described varieties of English, respectively British and American. Moreover, not even the *BACP* appears to record any of them—a fact which testifies to the relevance of Forde's (1987) accomplishment. Figure 1 shows a map of Barbados where the locations embedded in the toponymic idioms analyzed are indicated (Note 5).
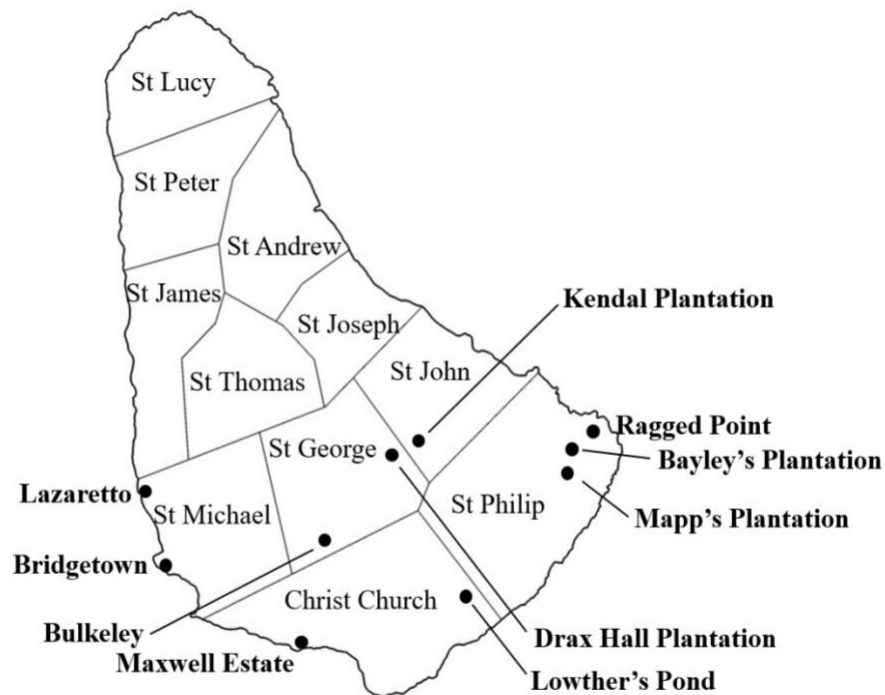
Figure 1. Location of places embedded in Bajan toponyms phrasemes

Evidence that all these toponymic phrasemes belong to Bajan, as opposed to Barbadian English, is granted by their non-standard morphological and orthographic features. Presumably more widely used in oral rather than in written communication—a feature that virtually all proverbs share, they include in their written renderings non-standard inflectional and spelling patterns which do not belong to English. Both the morphological and orthographic singularities of each idiom are highlighted in the following subsections. In this regard, it is helpful to keep in mind Sebba's (2007, p. 83) remark on the matter: "[i]f particular orthographic practices can be iconic of nations and ideologies, then getting rid of those orthographic practices may come to be seen as a part of rejecting colonialism and unwanted or imposed ideologies.".

Although many spelling standards for creoles have been devised worldwide over the past century, if attention is paid to the proposed writing systems for the English-lexified creoles of the Caribbean, such as those regarding Sranan Tongo, spoken in Suriname, and Jamaican Creole, also known as Patwa (Sebba, 2007, p. xix), their use is somewhat limited to official documents or academic publications. As for the former, the first official orthography for Sranan, closer to Dutch spelling standards, was developed in 1960, whereas a second official writing system, more akin to English orthographic conventions, was adopted in 1986 (see Sebba, 2007, p. 87-91). As for the latter, a standardized creole spelling model for Patwa was conceived in the 1960s by Cassidy and Le Page (1967/1980) in their *Dictionary of Jamaican English* (see Sebba, 2007, 119), while a revised version, presently known as the Cassidy-JLU (Jamaican Language Unit) writing system, was devised in 2002 (see Bogle, 2020, p. 26). Therefore, it goes without saying that the Bajan entries belonging to the collections surveyed—both amateur and professional—do not necessarily comply with the above-mentioned standards, only adopted in Suriname and Jamaica respectively. Regardless, the way in which entries are spelled still manages to mirror the peculiarities of Bajan speech and bring them into prominence. All in all, being contingent on the choices made by each compiler,

the adopted spellings may vary, at least to some extent, from collection to collection (see Hellinger, 1986, pp. 58, 62).

On a final note, before proceeding with the qualitative analysis of each toponymic phraseme considered, it is paramount to heed Forde's (1987, p. x) warning:

"In some instances there is an appearance of originality in the specific use of local names, as in "You can' tek down Drax Hall to put up Kendal," and "De Lazaretto dog dead at Ragged Point Lighthouse," but these could well have been adapted from other territories, with local names substituted, a fact which is difficult to prove or disprove in view of the limited research done in this area."

Accordingly, in the following subsections, the cultural evaluation conveyed by each phraseme—by definition, more semantically pregnant than any single lexical item *per se* (see Dobrovol'skij & Piirainen, 2022, p. 34)—is decoded through the etymological description of its toponymic constituent (see Dobrovol'skij & Piirainen, 2022, pp. 105, 317-320).

6.1 All me labour gone in Maxwell Pond *and* All o' me labour gone in Lowther's Pond

As for the morphological features of both sayings, the use of *me* instead of *my* and the absence of the auxiliary *has* before the head verb *gone* stand out as non-standard. From an orthographic perspective, *o'*, followed by an apostrophe, appears instead of *of* in the latter proverb. It is also interesting to notice the British spelling of *labour*, which indeed confirms that the creole of Barbados is based on varieties of English stemming from the British Isles.

The meaning of these two proverbs including Barbadian toponyms, namely *Maxwell Pond* and *Lowther's Pond*, may be condensed into 'someone's efforts or money have been wasted'. A similar version of the former, that is *All his money gone in Maxwell pond*, appears under the entry *Maxwell pond* in Collymore (2005, p. 67), who confirms that this is "a saying denoting that someone's money has been lost in some investment". Likewise, a similar meaning is provided by Richard Allsopp (2003) in the *Dictionary of Caribbean English Usage* (*DCEU*) under the entry *Maxwell pond*, which reads as follows:

"**Max·well pond** *n phr* (Bdos) PHRASE **labour / money gone in Maxwell Pond** *id phr* (Bdos) [*AF–Joc*] Effort or money has been lost, wasted in some unfortunate investment. […] **[**The phr is historically associated with the former Maxwell (Sugar) Estate in the PARISH of Christ Church in So. Bdos, though no pond or incident has been identified**]**"

However, in partial contrast to the last statement made by Allsopp in the *DCEU* definition, namely that "no pond or incident has been identified", it is the same Forde (1987, p. 40) who explains that "[a]t Maxwell Pond one day, a frog hunter dropped his half-full bag, forgetting the mouth was open, and started hunting. All his frogs escaped into the pond. He said, "All me labour gone in Maxwell Pond." All his efforts have been wasted.". The plausible reason why frogs used to be considered so precious and were hence hunted relies on the fact that they represented a valid natural alternative to eliminate insects long before pesticides. All in all, the inexistence of a pond by that name is confirmed by *Green's Dictionary of Slang*, which, under the entry *Maxwell Pond*, reads as follows: "proper name of the *Maxwell* (Sugar) Estate in Barbados. No actual pond, however, has ever been traced". *Green's Dictionary of Slang* also acknowledges the presence of the phrases *money gone in Maxwell Pond* and *labour gone in Maxwell Pond* in West Indian English speech, specifying that they are "used to describe money or effort that has been wasted or 'thrown away'". It must be added that the variant *All yuh labour gone in Maxwell pond* is also recorded in Blackman (1985, p. 1). The *Legacies of British Slavery* (*LBS*) database lists Maxwells as a Barbadian plantation possibly established in 1745 and owned by Thomas Maxwell until 1751.

As for the variant *All o' me labour gone in Lowther's Pond*, it seems to have a different etymology as it includes a biblical reference to Ecclesiastes 11 (John 10: 22-42), where Solomon states "Cast your bread upon the waters, for you will find it after many days.". Indeed, as Forde (1987, p. 40) reports, "[a] bread-seller took this advice literally, threw a tray of bread into Lowther's Pond, and visited the pond daily to see it return.". Although there is no present sign confirming the existence of a pond in the area, it was probably once located in the premises of Lowther's estate. As explained in the *LBS* database, Lowther's, a 214-acre plantation, was first established by Edmond Lewis in 1674. The plantation was later inherited through marriage by Robert Lowther, an English landowner who served as governor of Barbados from 1711 to 1720. In 1756, it passed into the ownership of his first son, James Lowther.

6.2 De Lazaretto dog dead at Ragged Point Lighthouse

From a morphological point of view, the copular verb *is/was*—and possibly the head verb *found*—is omitted between *dog* and *dead*. The typically creole spelling features of this adage are limited to the use of the article *de* instead of *the*.

As for its semantics, the toponymic proverb *De Lazaretto dog dead at Ragged Point Lighthouse* exploits two Barbadian place names corresponding to different locations. The Lazaretto, a building adjoining Batts Rock beach

in St Michael, almost five kilometers north of Bridgetown, the capital city, stands on the west coast, facing the Caribbean Sea. Ragged Point, in St Philip, is the most easterly tip of Barbados overlooking the Atlantic Ocean (see Furiassi, 2022, p. 100). The distance of about twenty-four kilometers between these two topographic landmarks—truly remarkable, especially considering that it covers the entire width of the island—is useful in explaining the meaning of the proverb, as indicated by Forde (1987, p. 2): "[d]o not expect people to put geographical limits on their movements (Also used in reference to gallivanting men.)".

The entry *lazaretto Dog Dead in St. Philip*, without the capitalization of the initial <l> in *lazaretto* and where *St. Philip*, the parish where the lighthouse is located, is mentioned instead of *Ragged Point*, is found in Collymore (2005, p. 60) and accompanied by the following comment: "[t]his quaint expression implies that some philandering husband has met his death while visiting his lady love far from home". Furthermore, the variant *De Lazaretto dog die at de lighthouse*—with no mention of either Ragged Point or St Philip—is recorded in Blackman (1985, p. 2), where, unfortunately, neither the meaning nor any other additional information is provided.

The Lazaretto, formerly the site of a British fort and battery, was converted into an asylum to treat victims of leprosy around 1863. At present, the Lazaretto complex hosts the offices of various public organizations, including the National Archives (see Forde, 1987, p. 2). The lighthouse at Ragged Point was built in 1875. After being long abandoned and temporarily extinguished between 2007 and 2011, it was eventually restored as a navigational aid and has become a tourist attraction since 2018 (see Rowlett, 2023).

6.3 Dis place hot as Mapp's Millyard *and* Dis place hot as Bayley's Millyard

As far as the morphology of this simile is concerned, the absence of the copula *is* between *place* and *hot* is worth considering. The creole nature of the spelling conventions applied to both entries is apparent in the use of *Dis* instead of *this*. Semantically, both *Dis place hot as Mapp's millyard* and its variant, *Dis place hot as Bayley's millyard*, appear to denote 'a place that is extremely hot' (see Furiassi, 2022, p. 99).

According to Forde (1987, pp. 37-38), the idiomatic reference to hotness may be connected to the uprising led by an African-born slave named Bussa, which took place in St Philip between April 14th and April 16th, 1816 (see Beckles, 2006, pp. 108-116). Indeed, during Bussa's revolt, the fire set by the insurgents to Bayley's plantation and its estate immediately extended to the adjoining Mapp's plantation. The simile *as hot as Mapp's mill-yard* is also recorded under the entry *Mapp's mill-yard* in Collymore's (2005) *Barbadian Dialect*. Even though he has "not been able to discover its origin", Collymore (2005, p. 66) admits that "the simile […] still lingers on among the older generation".

As per the *LBS* database, Mapps plantation—spelled without the apostrophe indicating possession—was owned by the Mapp family from at least 1720. As of more recent times, Forde (1987, p. 38) explains that Graham Wilkes rented the estate in 1957 and three years later, in 1960, founded on the premises the now defunct Mapp's College, a private secondary school. Besides, Bayley's plantation was established between 1719 and 1738; by 1765 it was in the possession of Joseph Bayley. As for the present, Guyana-born singer Eddy Grant acquired Bayley's plantation, where he currently resides, in 1981. He restored the main plantation house and converted the stables into a recording studio (see Forde, 1987, p. 38).

6.4 Dem does work at Idle Hall

The creole traits of the saying *Dem does work at Idle Hall*, whose meaning is "[t]hey are unemployed" (Forde, 1987, p. 41), are apparent. Morphologically, the auxiliary *does*, which substitutes *do* and would be normally used in (Standard) English to remark emphasis, is applied here to stress habituality; *Dem*, the orthographic rendering of *Them*, indicates the subject even though it appears in the objective case. To be able to explore the semantic implication of *Dem does work at Idle Hall*, readers must resort to the *DCEU*: *work at Idle Hall*, found within the entry *Idle Hall (Estate)*, means "[t]o be unemployed (and not too concerned about finding a job).". A similar saying, *To work at Walker's and get paid at Idlehall*, is listed in Blackman (1985, p. 11). Even in this case, the *DCEU* proves useful in discerning its hidden sense: the verb phrase *work(ing) at Walkers (& Co)*, present under the entry *Walkers & Co*, means "[b]e(ing) unemployed". According to the data provided in the *DCEU*, both entries are marked as "Bdos" and should hence be considered prototypically Bajan. However, it is worth pointing out that both *Walkers* and *Idle Hall* are made-up names of imaginary places which, quite straightforwardly, hint at a state of (permanent) inertia. In fact, a partially different explanation is provided by Collymore (2005, p. 114) under the entry *Walker's*, which reads as follows: "to be employed at *Walker's* (the name of a plantaton) is a euphemism signifying that the person referred to is walking around seeking emplyment".

6.5 Ent she hard! She just like Bulkeley iron man

Considering the morphological features of this simile, the deletion of the copulative verb *is* between *She* and *just* is an apparent creole trait. As regards spelling, *Ent* is used to render *Isn't*. Quite curiously, the toponymic simile considered is included in a made-up example intended to mimic a conversation taking place in Bajan. All in all, *To be hard like Bulkeley iron man* is apparently intended to address an individual who is "hard (skinny) as iron"

(Forde, 1987, p. 38). Although no Barbadian or Caribbean connotation of *hard* as 'thin' is present in the *DCEU*, the meaning provided by Forde is presumably based on the resemblance between said person and the material used to build the largest steam-operated mill of any sugar factory in Barbados, installed at Bulkeley, in the parish of St George, in the 1920s (see Forde, 1987, p. 38). The factory is named after William Bulkeley, who, as claimed by the *LBS* database, was "listed as a landowner in 1638" and whose "plantation remained in the hands of the Bulkeley family until 1724". At present, the Bulkeley Sugar Factory still functions as one of the oldest continuously operating sugar factories on the island (see Forde, 1987, p. 39).

6.6 He gine promise you Drax Hall and Kendal

As for the morphology of this proverb, the omission of the copula *is* between *He* and *gine* is evident. As far as its spelling is concerned, *gine* is used instead of *going to*. *To promise Drax Hall and Kendal*, meaning 'to make hollow promises', corresponds to *promise (someone) the stars/moon/earth/world*, that is "to promise (someone) that one will do or give something great or wonderful even though it is not possible" (*Merriam-Webster's Collegiate Dictionary*). The reason for the inclusion of these two place names in this proverb is logically explained by the fact that "[t]hose two plantations were at one time the two largest in Barbados" (Forde, 1987, p. 42). It is worth mentioning that a similar entry, *He'll promise you Drax Hall and Kendal*, is found in Blackman (1985, p. 5), who only adds that the place names mentioned in the saying are "two estates in Barbados".

As maintained by the *LBS* database, Drax Hall, located in St George, was built by William and James Drax in the 1650s. Colonel James Drax, who also owned a plantation by the same name in St Ann, Jamaica, is known to have introduced sugar cane cultivation to Barbados as early as 1642, not long before the outset of the "Sugar Revolution" (O'Callaghan, 2000, p. 93; Beckles, 2006, p. 27), which took place in 1644 and converted the economy of the island from tobacco to sugar production. Nowadays, the Drax family still retains the possession of the estate, which remains a sugar plantation. Drax Hall is one of the only two intact Jacobean houses remaining in Barbados—the other one being St Nicholas Abbey in St Peter (see Forde, 1987, p. 22). On the contrary, little is known about Kendal, except that, in line with the information provided by the *LBS* database, it was established by Joshua Steele in 1757 and was then known as Kendals—with the addition of the final <s> in the name.

6.7 You don't tek down Drax Hall to fix Kendal *and* You can' tek down Drax Hall to put up Kendal

The only non-standard features appearing in both these proverbs relate to spelling: *tek* appears instead of *take* and *can*, followed by an apostrophe, is used instead of *can't*. The formal differences between *You don't tek down Drax Hall to fix Kendal* and *You can' tek down Drax Hall to put up Kendal* lie in the use of the modal *can'* and the presence of *put up*, instead of *fix*, in the latter proverb. Their semantic interpretation is related to the concept of Drax Hall and Kendal being the two wealthiest estates in Barbados (see Section 6.6). Indeed, as detailed by Forde (1987, p. 22), "Drax Hall plantation, with 881 acres, was and still is the largest single plantation in Barbados; Kendal, in St. John, is the second largest, with 718 acres". Forde (1987, p. 22) adds that the meaning of both proverbs can be explained as follows: "[i]t does not make sense destroying something of value in order to make right something of equal or less value".

7. Conclusion

In general, this piece of research has shown how "dictionaries […] reflect […] the relation between language and culture" (Gouws, 2020, p. 3). Transitively, the present investigation—however limited—has proven that the cultural aspects of Barbadian society are mirrored in its paremiographic heritage. Indeed, the idiomatic toponymic phrasemes analyzed all "contain figurative elements with cultural implications" (Cotta Ramusino & Mollica, 2020, p. 4) and act as carriers of "culture-boundness" (Sabban, 2007, p. 590; 2008, p. 232).

Toponymic phrasemes were selected by relying on the metalexicographic analysis of Forde's (1987) glossary, a personal, passionate effort which, by narrating Bajan culture, gives voice to the Barbadian speech community and bears relevance to the English-based creole it speaks. Forde managed to bring to the forefront aspects of local heritage that would have likely been lost otherwise. His honest cataloguing of territorial paremiology successfully validates the linguistic tradition of Barbadians, hence preserving it for both present and succeeding generations. This tribute to Barbadianness certainly merits to be recognized as a landmark in the paremiographic landscape of the island.

From a quantitative perspective, the paremiological inventory of Bajan emerges as bountiful not only thanks to the inclusion of brand-new proverbs and proverb-like phrases arising from long-established English lexical material, but also thorough the "resemanticization" (Paganoni, 2007, p. 187) of English-derived toponyms which, despite pertaining to the harsh colonial history of Barbados, have been recontextualized and semantically redetermined. Moreover, from a qualitative viewpoint, this article has shown how the uniqueness of Barbadian place names gives birth to equally distinctive Bajan idiomatic phrasemes that can only be unearthed from the communicative practices of islanders.

Unsurprisingly, most of the toponyms detected, namely *Bayley*, *Bulkeley*, *Drax*, *Kendal*, *Lowther*, *Mapp* and *Maxwell*, denote former sugar-cane plantations established between the 17th and the 18th century and the attendant estates around which slaves were forced to work. In virtually all cases, said toponyms clearly derive from their respective anthroponyms—equally endowed with connotational nuances (see Szerszunowicz, 2012, p. 297; Dobrovol'skij & Piirainen, 2022, p. 318)—identifying the last names of early British owners. The messages carried by the toponymic Bajanisms described draw attention to the archetypal weltanschauung of Barbadians, which breaks away from the English that had been superimposed over nearly 400 years of British dominion. Although the collective recollection of the slavery era might be irremovable, the above findings bear witness to the fact that Bajans have long ceased to live (and speak) under the enduring shadow of their colonial past.

Before concluding, two desiderata seem in order. To begin with, an aspect that should be considered for further research is the verification whether the idiomatic toponymic phrasemes discussed are actually (still) used by contemporary speakers of Bajan or, being the sole possession of the elderly, they are inexorably doomed to extinction. Then, it would be equally worthwhile to retrieve all the lexical and phraseological coinages—not just the proverbs—inspired by Barbadian toponyms with the goal of systematizing them. For the sake of exhaustiveness, this survey should be performed by investigating the toponymic entries encountered in all the Bajan (and Barbadian English) lexicographic and paremiographic products available, regardless of whether their approach is scholarly or essentially amateur.

References

Alleyne, M. C. (1971). Acculturation and the Cultural Matrix of Creolization. In D. Hymes (Ed.), *Pidginization and Creolization of Languages* (pp. 169-186). Cambridge, England: Cambridge University Press.

Allsopp, J. (2009). Dictionaries of Caribbean English. In A. P. Cowie (Ed.), *The Oxford History of English Lexicography* (Vol. 1, pp. 353-377). Oxford, England: Oxford University Press.

Allsopp, J., & Furiassi, C. (2020). Caribbean English Phraseology in the *Dictionary of Caribbean English Usage*: Reflections of an African Worldview. *TEXTUS: English Studies in Italy, XXXIII*(1), 107-125.

Allsopp, R. (2003). *Dictionary of Caribbean English Usage* (2nd imp.). Mona, Jamaica: University of the West Indies Press. https://doi.org/10.37234/TCNDYYBE

Allsopp, R. (2004). *A Book of Afric Caribbean Proverbs*. Kingston, Jamaica: Arawak Publications.

Allsopp, R. (2000/2006). The Afrogenesis of Caribbean Creole Proverbs. *Society for Caribbean Linguistics Occasional Papers, 34*, 36-52.

Beckles, H. McD. (2006). *A History of Barbados* (2nd ed.). Cambridge, England: Cambridge University Press.

Blackman, M. (1985). *Bajan Proverbs* (2nd ed.). Montreal, Canada: Margot Blackman.

Bogle, D. (2020). *The Transatlantic Culture Trade: Caribbean Creole Proverbs from Africa, Europe, and the Caribbean*. New York, NY: Peter Lang. https://doi.org/10.3726/b14121

Blake, R. (2004). Bajan: Phonology. In E. W. Schneider, K. Burridge, B. Kortmann, R. Mesthrie, & C. Upton (Eds.), *A Handbook of Varieties of English* (Vol. 1, pp. 501-507). Berlin, Germany, & New York, NY: Mouton de Gruyter.

Bredis, M., & Lomakina, O. (2019). Toponyms in Proverbs: A Case Study of Russian, Latvian, Lithuanian, Polish, German, French, English, Finnish, and Tajik Proverbs. *Proverbium: Yearbook of International Proverb Scholarship, 36*(1), 75-92.

Brooks, S. (1985). Preface. In M. Blackman, *Bajan Proverbs* (2nd ed., pp. vi-ix.). Montreal, Canada: Margot Blackman.

Cassidy, F. G., & Le Page, R. B. (1967/1980). *Dictionary of Jamaican English*. Cambridge, England: Cambridge University Press.

Collymore, F. A. (1966). Barbadian Proverbs. *Bim, 11*(43), 158-163.

Collymore, F. A. (2005). *Barbadian Dialect: Notes for a Glossary of Words and Phrases of Barbadian Dialect* (7th ed.). Bridgetown, Barbados: Cave Shepherd & Co.

Cotta Ramusino, P., & Mollica, F. (2020). Contrastive Phraseology: Preliminary Remarks. In P. Cotta Ramusino, & F. Mollica (Eds.), *Contrastive Phraseology: Languages and Cultures in Comparison* (pp. 1-10). Newcastle upon Tyne, England: Cambridge Scholars Publishing.

Cowie, A. P. (1998). Introduction. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications* (pp. 1-20). Oxford, England: Oxford University Press. https://doi.org/10.1093/oso/9780198294252.003.0001

Cowie, A. P., Mackin, R., & McCaig, I. R. (1993). General Introduction. In *Oxford Dictionary of English Idioms* (pp. x-xvii). Oxford, England: Oxford University Press.

Dobrovol'skij, D., & Piirainen, E. (2022). *Figurative Language: Cross-cultural and Cross-linguistic Perspectives* (2nd ed.). Berlin, Germany, & Boston, MA: De Gruyter Mouton. https://doi.org/10.1515/9783110702538

Dueck, D. (2004). 'Bird's Milk in Samos': Strabo's Use of Geographical Proverbs and Proverbial Expressions. *Scripta Classica Israelica, XXIII*, 41-56.

Ethnologue (25th ed.). (2022). Dallas, TX: SIL International. Retrieved December 22, 2023, from http://www.ethnologue.com

Fenigsen, J. (1999). "A Broke-up Mirror": Representing Bajan in Print. *Cultural Anthropology, 14*(1), 61-87. https://doi.org/10.1525/can.1999.14.1.61

Fenigsen, J. (2003). Language Ideologies in Barbados: Processes and Paradigms. *Pragmatics, 13*(4), 457-482. https://doi.org/10.1075/prag.13.4.01fen

Fenigsen, J. (2007). From Apartheid to Incorporation: The Emergence and Transformations of Modern Language Community in Barbados, West Indies. *Pragmatics 17*(2), 231-261. https://doi.org/10.1075/prag.17.2.02fen

Fenigsen, J. (2011). "Flying at Half-mast"? Voices, Genres, and Orthographies in Barbadian Creole. In L. Hinrichs, & J. T. Farquharson (Eds.), *Variation in the Caribbean: From Creole Continua to Individual Agency* (pp. 107-132). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/cll.37.08fen

Finegan, E. (2020). Editorial. *DICTIONARIES: Journal of the Dictionary Society of North America, 41*(2), ix–xii. https://doi.org/10.1353/dic.2020.0017

Forde, G. A. (1987). *De Mortar-pestle: A Collection of Barbadian Proverbs*. St James, Barbados: The National Cultural Foundation.

Furiassi, C. (2022). Detecting Bajan Phraseology: A Metalexicographic Analysis. In G. Corpas Pastor, & R. Mitkov (Eds.), *Computational and Corpus-Based Phraseology* (pp. 90-104). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-031-15925-1_7

Furiassi, C. (2023). Barbadian English Idioms: Challenging Linguistic Norms in a Diasporic Context. *Testi e Linguaggi, 17*, 46-65.

Furiassi, C. (in press). A History of Amateur Anglophone Lexicography and Lexicographers in Barbados. *DICTIONARIES: Journal of the Dictionary Society of North America*, *46*(1).

Gibson, K. (1996). Observations on *have* and *had* in Barbadian and Guyanese Mesolectal English. In P. Christie (Ed.), Caribbean Language Issues: Old and New. Papers in Honour of Professor Mervyn Alleyne on the Occasion of His Sixtieth Birthday (pp. 41-47). Kingston, Jamaica: The Press University of the West Indies.

Gouws, R. H. (2020). Metalexicography, Dictionaries and Culture. *Lexicographica, 36*, 3-9. https://doi.org/10.1515/lex-2020-0001

Green's Dictionary of Slang. (2017). London, England: Chambers. Retrieved December 22, 2023, from https://greensdictofslang.com

Gries, S. Th. (2008). Phraseology and Linguistic Theory: A Brief Summary. In S. Granger, & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 3-25). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/z.139.06gri

Haynes-Knight, K.-A., Evans, K., & Winters-Evans, T. (2015). Wuh Allsopp tink she talking 'bout? Bajan vs Standard English as Mother Tongue. In N. Faraclas, R. Severing, C. Weijer, E. Echteld, & W. Rutgers (Eds.), *Envisioning the Greater Caribbean: Transgressing Geographical and Disciplinary Boundaries* (pp. 311-319). Willemstad, Curaçao: Fundashon pa Planifikashon di Idioma, & Universidat di Kòrsou.

Hellinger, M. (1986). On Writing English-related Creoles in the Caribbean. In M. Görlach, & J. A. Holm (Eds.), *Focus on the Caribbean* (pp. 53-70). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/veaw.g8.04hel

Ibekwe, P. (1998). *Wit & Wisdom of Africa: Proverbs from Africa & the Caribbean*. Oxford, England: WorldView Publishing.

Ibekwe, P. (2009). *The Little Book of African Wisdom*. Oxford, England: New Internationalist.

Kispál, T. (2015). Paremiography: Proverb Collections. In H. Hrisztova-Gotthardt, & M. A. Varga (Eds.), *Introduction to Paremiology: A Comprehensive Guide to Proverb Studies* (pp. 229-242). Berlin, Germany, & Boston, MA: Mouton de Gruyter.

Klein, J. R., & Lamiroy, B. (2016). Le figement: Unité et diversité. Collocations, expressions figées, phrases situationnelles, proverbs. *L'Information grammaticale, 148*, 15-20.

Lambert, J. (2020). Lexicography and World Englishes. In D. Schreier, M. Hundt, & E. W. Schneider (Eds.), *The Cambridge Handbook of World Englishes* (Part III, pp. 408-435). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/9781108349406.018

Legacies of British Slavery. (2023). London, England: University College London. Retrieved December 22, 2023, from http://wwwdepts-live.ucl.ac.uk/lbs/estates

Marshall, T. G. (1987). Foreword. In G. A. Forde, *De Mortar-pestle: A Collection of Barbadian Proverbs* (p. v). St James, Barbados: The National Cultural Foundation.

McArthur, T., Lam-McArthur, J., & Fontaine, L. (Eds.). (2018). *The Oxford Companion to the English Language* (2nd ed.). Oxford, England: Oxford University Press. https://doi.org/10.1093/acref/9780199661282.001.0001

Mel'čuk, I. (1995). Phrasemes in Language and Phraseology in Linguistics. In M. Everaert, E. J. van der Linden, A. Schenk, & R. Schreuder (Eds.), *Idioms: Structural and Psychological Perspectives* (pp. 167-232). Mahwah, NJ: Lawrence Erlbaum Associates.

Mel'čuk, I. (2012). Phraseology in the Language, in the Dictionary, and in the Computer. *Yearbook of Phraseology, 3*, 31-56. https://doi.org/10.1515/phras-2012-0003

Merriam-Webster's Collegiate Dictionary (11th ed.). (2019). Springfield, MA: Merriam-Webster. Retrieved December 22, 2023, from https://unabridged.merriam-webster.com

Merriam-Webster's Unabridged Dictionary (3rd ed.). (2022). Springfield, MA: Merriam-Webster. Retrieved December 22, 2023, from https://unabridged.merriam-webster.com

Mieder, W. (1996). Proverbs. In J. H. Brunvand (Ed.), *American Folklore: An Encyclopedia* (pp. 597-621). New York, NY: Garland.

Mieder, W. (2012). *Proverbs: A Handbook* (2nd ed.). New York, NY: Peter Lang.

Norrick, N. R. (2015). Subject Area, Terminology, Proverb Definitions, Proverb Features. In H. Hrisztova-Gotthardt, & M. A. Varga (Eds.), *Introduction to Paremiology: A Comprehensive Guide to Proverb Studies* (pp. 7-27). Berlin, Germany, & Boston, MA: Mouton de Gruyter.

Nuccorini, S. (2006). In Search of 'Phraseologies': Discovering Divergences in the Use of English and Italian 'True Friends'. *European Journal of English Studies, 10*(1), 33-47. https://doi.org/10.1080/13825570600590879

Nuccorini, S. (2016). Phraseology in Time: On the Innovative Treatment of Word Combinations in Specialised Nineteenth-Century Bilingual Dictionaries. *Language & History, 59*(1), 48-62. https://doi.org/10.1080/17597536.2016.1189665

O'Callaghan, S. (2000). *To Hell or Barbados: The Ethnic Cleansing of Ireland*. Dublin, Ireland: Brandon.

Oxford English Dictionary Online. (2023). Oxford, England: Oxford University Press. Retrieved December 22, 2023, from https://www.oed.com

Paganoni M. C. (2007). Recontextualizing Language: Indian Activists and the Recasting of English. In G. Garzone, & C. Ilie (Eds.), *The Use of English in Institutional and Business Settings: An Intercultural Perspective* (pp. 177-196). Bern: Peter Lang.

Piirainen, E. (2008). Figurative Phraseology and Culture. In S. Granger, & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 207-228). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/z.139.20pii

Pulcini, V., Furiassi, C., & Rodríguez González, F. (2012). The Lexical Influence of English on European Languages: From Words to Phraseology. In C. Furiassi, V. Pulcini, & F. Rodríguez González (Eds.), *The Anglicization of European Lexis* (pp. 1-24). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/z.174.03pul

Rickford, J. R. (1992). The Creole Residue in Barbados. In N. Doane, J. Hall, & D. Ringler (Eds.), *Old English and New: Essays in Language and Linguistics in Honor of Frederic G. Cassidy* (pp. 183-201). New York, NY: Garland.

Rickford, J. R., & Blake, R. (1990). Copula Contraction and Absence in Barbadian English, Samaná English and Vernacular Black English. In D. J. Costa (Ed.), *Proceedings of the Sixteenth Annual Meeting of the Berkeley*

Linguistics Society (pp. 257-268). Berkeley, CA: Berkeley Linguistics Society. https://doi.org/10.3765/bls.v16i0.1683

Rowlett, R. (2023). *Lighthouses of Barbados*. Chapel Hill, NC: The University of North Carolina. Retrieved December 22, 2023, from https://www.ibiblio.org/lighthouse/brb.htm

Sabban, A. (2007). Culture-boundness and Problems of Cross-cultural Phraseology. In H. Burger, D. Dobrovol'skij, P. Kühn, & N. R. Norrick (Eds.), *Phraseology: An International Handbook of Contemporary Research* (Vol. 1, pp. 590-605). Berlin, Germany, & Boston, MA: De Gruyter Mouton. https://doi.org/10.1515/9783110171013.590

Sabban, A. (2008). Critical Observations on the Culture-boundness of Phraseology. In S. Granger, & F. Meunier (Eds.), *Phraseology: An Interdisciplinary Perspective* (pp. 229-241). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/z.139.21sab

Schneider, E. W. (2007). *Postcolonial English: Varieties around the World*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511618901

Sebba, M. (2007). *Spelling and Society: The Culture and Politics of Orthography around the World*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/CBO9780511486739

Stuka, C. (2023). The Americanization of Barbadian English. *World Englishes, 42*(1), 91-114. https://doi.org/10.1111/weng.12611

Szerszunowicz, J. (2009). Some Remarks on the Evaluative Connotations of Toponymic Idioms in a Contrastive Perspective. In R. Corrigan, E. A. Moravcsik, H. Ouali, & K. Wheatley (Eds.), *Formulaic Language* (Vol. 1, pp. 171-184). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/tsl.82.08som

Szerszunowicz, J. (2012). On the Evaluative Connotations of Anthroponymic Idioms in a Contrastive Perspective (Based on English and Italian). *Białostockie Archiwum Językowe, 12*, 293-314. https://doi.org/10.15290/baj.2012.12.18

Van Herk, G. (2003). Barbadian Lects: Beyond Meso. In M. Aceto, & J. P. Williams (Eds.), *Contact Englishes of the Eastern Caribbean* (pp. 241-264). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/veaw.g30.14her

Winer, L. (2006). Caribbean Lexicography. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., Vol. 2, pp. 204-205). Amsterdam, The Netherlands: Elsevier. https://doi.org/10.1016/B0-08-044854-2/05151-8

Winford, D. (2000). "Intermediate" Creoles and Degrees of Change in Creole Formation: The Case of Bajan. In I. Neumann-Holzschuh, & E. W. Schneider (Eds.), Degrees of Restructuring in Creole Languages (pp. 215-246). Amsterdam, The Netherlands, & Philadelphia, PA: John Benjamins. https://doi.org/10.1075/cll.22.13win

Winford, D. (2003). *An Introduction to Contact Linguistics*. Oxford, England: Blackwell.

Notes

Note 1. According to the *Merriam-Webster's Unabridged Dictionary*, a *Wellerism*—an eponym first used in 1838 and stemming from Samuel Weller, a character in the *Pickwick Papers* by the English novelist Charles Dickens—is "an expression of comparison comprising a usually well-known quotation followed by a facetious sequel".

Note 2. The definition of proverbs sensu stricto provided by Mieder (1996, p. 597) reads as follows: "[…] short, generally known sentences of the folk that contain wisdom, truths, morals, and traditional views in a metaphorical, fixed, and memorizable form and that are handed down from generation to generation". Alternatively, Norrick (2015, p. 14) defines proverbs as "recurrent, pithy, often formulaic and/or figurative, fairly stable and generally recognizable units used to form a complete utterance, make a complete conversational contribution and/or to perform a speech act in a speech event. This definition differentiates them from non-sentential items like proverbial phrases, idioms, binomials etc.".

Note 3. In Kispál's (2015, p. 240) words, "[c]ollections containing mainly proverbs, except of few, are often mere lists without any information on their meaning and usage. Proverbs are ordered alphabetically mostly by keywords. In thematically organized proverb collections, the starting points are alphabetically ordered topics to which proverbs are assigned".

Note 4. A preliminary study on Bajan toponyms extracted from a wide range of lexicographic—though not strictly paremiographic—sources has been recently carried out by Furiassi (2022).

Note 5. Image sourced from Wikimedia Commons https://commons.wikimedia.org/wiki/File:Barbados_parishes_blank.png. Adapted and shared in compliance with Creative Commons license CC BY 4.0 DEED https://creativecommons.org/licenses/by/4.0/deed.

Copyrights

Linguacultural representations in specialized migration discourse: a lexicographic perspective

Silvia Sperti[1]

[1] Department of Foreign Languages, Literatures and Cultures, 'Roma Tre' University, Rome, Italy

Correspondence: Silvia Sperti, Department of Foreign Languages, Literatures and Cultures, 'Roma Tre' University, Via Ostiense 234-236, 00146 Rome, Italy. E-mail: silvia.sperti@uniroma3.it

Abstract

In the last two decades increasing movements of people across countries, due to economic and social reasons, have produced high levels of exchanges among speakers of different languages where English is used globally as a lingua-franca (ELF). Moreover, migration flows across nation states (especially from non-Western countries to Western ones) have encouraged the movement of people, mainly of African and Asian origin, from students to skilled workers, who are often involved in English-mediated interactions where migrants' native linguacultural background inevitably connects to the language spoken by the host community (e.g. in European countries), and at the same time shapes the use of English as a global means of interaction (Canagarajah, 2013). ELF cross-cultural interactions and translingual processes, naturally occurring in intercultural settings, are particularly remarkable, since they plainly show how ELF speakers, engaged in intercultural interactions, differently appropriate the English language, according to their own native linguacultural patterns, and to specific pragmalinguistic purposes and processes (Guido, 2012, 2018; Mauranen, 2018). This study will address the influence that lexical actualizations in authentic spoken encounters, as well as in written productions in specialized contexts have on the current role and function of English as an international language and which deserve coverage and consideration in lexicographic resources. Research studies on migration narratives, language mediation, cross-cultural conceptual representation and reception of traumatic events, where ELF lexical processes are often activated by the speakers involved, are particularly important to address the development of linguacultural representations that should be covered in dictionaries, lexicons and other lexicographic resources, especially online.

Keywords: lexical innovation, lexicography, migration discourse, ELF (English as a Lingua Franca), corpora.

1. Introduction

The emerging new linguistic landscapes appearing from migration encounters in Europe are particularly indicative of the impact that multilinguals and their multicultural backgrounds may have on the current research in applied linguistics. The investigation of intercultural interactions (e.g. Firth, 2009; Jenkins et al., 2011; Mauranen, 2012), involving not only English but a multilingual repertoire pertaining to each speaker, shows that they control and adjust to their interlocutors, engaging in the "strategic negotiation of the linguistic resources" that allow "the co-construction of understanding" (Seidlhofer, 2011: 198). Multilingual repertoires can be described as emergent linguistic and communicative practices that interlocutors naturally and spontaneously adopt to foster mutual understanding and to create a proactive interactional space. Research findings indicate that speakers in multilingual communicative contexts demonstrate attitudes and abilities to creatively exploit their linguacultural resources in ways that affect the lexical dimensions of the standardized forms of the language, especially English (Mauranen, 2012; Seidlhofer, 2011). "It is thus not so much uniformity of form, but communicative alignment, adaptation, local accommodation and attunement that would appear to underpin successful lingua franca interactions" (Firth, 2009: 163).

As a result, the focus is here on the influence that lexical actualizations in authentic spoken encounters, in migration contexts, as well as in written productions have on the current role and function of English as an international language which deserve attention and consideration from a lexicographic perspective.

2. Emerging multilingual landscapes: the role of migration

In the last three decades the intensification of social fragmentation processes ascribable to the recent tidal migration flows taking place almost all over the world, together with the growing diffusion of new technologies, social networks and multimodal communication, have contributed to creating new sociolinguistic environments while transforming the linguistic landscape of most cities (Dendrinos, 2012; Shohamy & Gorter, 2009).

We are now witnessing the highest levels of displacement on record; as shown by the latest data released by the United Nations High Commissioner for Refugees (UNHCR), at the end of 2022, 108.4 million around the world have been forced from home. Among them are refugees, even under the age of 18, displaced people, and asylum

seekers. There are also millions of stateless people who have been denied a nationality and access to basic rights such as education, healthcare, employment and freedom of movement. Europe has been a crossroads of human mobility since ancient times. Throughout history, the region has been a central part of global migration systems which its states helped to establish and shape. Europe also played a crucial role in developing a set of rules and norms regulating human mobility in the region. Within the last decades, European states have witnessed the mobility of high- and low skilled workers from Central and Eastern Europe to Western and Southern Europe as well as new waves of immigration from North and Central Africa, Latin America and Asia to Southern Europe. Meanwhile, the geopolitical conflicts in Ukraine, in the Middle East and in North Africa prompted an increase in the numbers of arrivals in Southern Europe of asylum seekers trying to reach Northern European destinations. In this perspective, the European Union (EU) has adopted various sets of rules and frameworks to manage legal migration flows for asylum seekers, highly skilled workers, students and researchers, seasonal workers, and family reunification. Regarding other migration flows, the EU has adopted common rules for processing asylum requests: first of all, the same procedure to relocate thousands of asylum seekers from Greece and Italy, and re-admission agreements for returning illegal migrants. Asylum is granted to people who are fleeing persecution, war or serious harm in their own country and, therefore, in need of international protection. In practice, anyway, the current system is still characterised by differing treatment of asylum-seekers and varying recognition rates amongst EU Member States. This divergence is what encourages secondary movements and is partly due to the fact that the current rules grant Member States a lot of discretion in how they apply the common EU rules. The EU now needs to put in place the tools to better manage migration flows. The overall objective is to move from a system which encourages uncontrolled or irregular migratory flows to one which provides safe pathways to the EU for third country nationals.

The migration of people moving for economic or political reasons from different countries or geographical areas in Asia, the Middle East and Africa as well as from South America, towards Europe, is still taking place in several countries, modifying the sociolinguistic features of the European cities, as well as determining a change in the societal conditions of their now multilingual and multicultural population. It is within these new scenarios that languages themselves are undergoing a unique transformative process, particularly expanding their lexis with new loan words and with multiethnolects (Cheshire et al., 2015), as well as extending traditional communicative functions and notions of contact language and interculturality (Baker, 2015, 2020), in order to sustain successful communication among speakers with different mother tongues, using either the country official language or a third language, that is very often English (Hoffman, 2020). In these new multilingual and multicultural contexts, mediation has become an essential measure aimed at reducing the distance between two (or more) poles of otherness, a notion officially introduced by the Council of Europe's Common European Framework of Reference (2020), where mediation acquires a crucial role as a new form of managing the interaction in language learning as well as in daily communicative situations.

Mediation has recently been more and more related to migration policies, socio-cultural inclusion, intercultural communication, language teaching, translation strategies and the use of English as a global language or as a lingua franca (henceforth, ELF) (Lopriore, 2015; Sperti, 2017). Mediation and mediation strategies are central in communication contexts where non-native speakers interact in environments where there is an increasing use of English as a lingua franca. Mediation emerges as a process activated in ELF communication, as it facilitates socialization and cooperation among participants who 'otherwise may not be able to participate' (Hynninen, 2011: 965). In European contexts, the flow of migrants within countries traditionally identified as monolingual and monocultural, has radically changed the linguistic landscapes giving place to multicultural and plurilinguistic scenarios where communication is inevitably 'intercultural' and mostly carried out in a lingua franca such as ELF. In this conceptual perspective, authentic processes of cross-cultural mediation, especially in migration contexts, are very often characterized by the employment of ELF lexical variations. The negotiation of meaning is carried out through the simultaneous action of different linguistic levels – namely linguistic, paralinguistic and extralinguistic ones – variously and creatively exploited by ELF users. Indeed, speakers reveal the use of different strategies of appropriation of the English language according to L1 linguacultural schemata and pragmalinguistic processes. In these multifaceted interactional dimensions, the exchange of speech acts affects the whole communicative process, from speakers' prosodic strategies to lexical and conversational dynamics (Sperti, 2017). The "mediator" is called to bridge gaps and overcome misunderstandings, to meet Kramsch's (1993) "third space" in which a speaker/learner might take some distance from his/her cultural norms to think critically and act as a social agent in two-pole interactions.

2.1 English as a Lingua Franca, language mediation and migration discourse

The spread of English as a global lingua franca and the attention to the related socio-linguistic phenomena is unquestionable as confirmed by the large amount of research about ELF in the last decades. As claimed by Mauranen (2018: 8), ELF is "English is spoken in situations with widely varying combinations of participants,

including first-language speakers of different varieties". In other words, it may be defined as "a contact language between speakers or speaker groups when at least one of them uses it as a second language" (Mauranen, 2018: 8).

Seen from this perspective, the use of English as a shared common language is particularly frequent in migration contexts where speakers from different socio-cultural and pragmalinguistic backgrounds exchange meaning for communicative purposes. In migration domains, mediating settings involve Western experts – namely legal advisors or officers, and mediators – and non-Western migrants (asylum-seekers and refugees included) interacting in professional contexts. As previously highlighted, in such communicative conditions ELF consists of variations developed from speakers' L1 structures transferred onto English. In other terms, ELF users act according to a process of authentication (Widdowson, 1979), shaping the English language by means of and according to their native pragmalinguistic conventions. The pragmalinguistic investigation of such interactions allows (i) to detect lexical processes and forms of appropriation across different socio-cultural backgrounds, and (ii) to define linguacultural representations emerging in ELF encounters, at times ending in communication failure or miscommunication (Guido, 2008; Sperti, 2017).

Mediation is an everyday activity occurring in public, educational, academic, and professional settings and in today's globalised world, very frequently, it is cross-linguistic. This is particularly true in migration contexts where both migrants and professionals who work with migrants often find themselves in situations in which they constantly need to intervene to create pluricultural space, facilitate communication, avoid misunderstandings and manage delicate situations. In this way speakers, involved at different levels, acquire and over time increase their intercultural awareness by means of linguacultural processes. Therefore, the investigation of language use and of mediation processes in these interactional situations has gained scientific interest in the last two decades, especially in the Mediterranean countries where migration flows are constant and their handling often very challenging (i.e. in Italy, Greece, Turkey and Spain). Mediation has been researched in the field of interlingual and intercultural communication in migration contexts (e.g. Guido, 2008, 2018; Sperti, 2017), where the use of English as a lingua franca has been explored with the aim of detecting hybridization strategies of reformulation aimed at "making ELF discourse conform to the immigrants' different native linguacultural backgrounds in order to protect the social identities of participants in unequal encounters, facilitate the mutual conveyance of their culturally-marked knowledge, foster successful intercultural communication through ELF, and finally promote the social inclusion of marginalized immigrants" (Guido, 2015:157).

At the same time, the use of ELF in mediating contexts is also characterized by forms of misunderstanding where non-Western migrants' native pragmalinguistic schemata collide with Western conformity to Standard-English models. In this unequal and unbalanced communicative infrastructure, successful mediation is challenged by speakers' pragmalinguistic behaviours that very frequently reveal conceptual gaps, cultural or ethical constraints, or cognitive and linguistic inaccessibility. Spontaneous cross-cultural interactions between asylum-seekers, language mediators and legal advisors are the basis for the research fieldwork of this paper whose aim is to provide a preliminary exploration, from a lexicographic perspective, of lexical variation in ELF (often associated to Italian lingua franca realizations), during cross-cultural exchanges that often involve the use of specialized spoken discourse in the context of legal counselling and assistance.

3. Research focus: from English variations to lexicographic divergences in representing migration discourse

Such theoretical grounds are thus at the basis of the research objectives aimed at enquiring into the use of lexical strategies applied by ELF speakers from different L1 backgrounds in immigration domains, accounting for (i) the influence of existing L1 correlates and transfers into ELF variations; (ii) the construction of meaning and understanding in cross-cultural interactions through lexical strategies applied to the negotiation of speakers' attitudes, and socio-cultural 'schemata'; (iii) processes of lexical creativity in language mediation activated in specialized immigration domains emerging from cross-cultural legal-bureaucratic and post-traumatic conversational settings. In this perspective and in order to highlight the complexity of a linguacultural dimension, that struggles to find an effective representation in lexicographic resources dealing with migration discourse, the emerging English variations are here explored in two different actualizations: the first one concerning the institutional and official terminology employing the EU variety of English, the so-called Euro-English; and the second one emerging from real practices of intercultural exchanges in professional contexts, such as the centres for legal advice or healthcare services, where English is commonly employed as a lingua franca.

3.1 ESP terminology in the EU institutions and Euro-English variations

In European institutions, English is the most widely spoken non-native language. It is particularly true after Brexit as suggested by Modiano in 2021, who argues that EU institutions should define and promote their own variety of English, complete with "punctuation, spelling, some grammar, and some vocabulary" (Note 1).

As matter of fact, Jenkins et al. (2001) underlined that lexical innovations eventually result in Euro-English: "through processes of discoursal nativization, wherein European expressions and conceptualisations that are

foreign to native-speaker varieties become valuable communicative tools, and fossilization, where 'nonstandard' structures become acceptable forms of language, as well as the existence of distinct European 'accents,' a new variety of English peculiar to the European experience is taking form. The lexical register of mainland European non-native speakers of English, steeped in standard English usage, is augmented by a number of terms which are peculiar to the European experience" (Jenkins et al., 2001: 12)

In the last two decades linguistic research on Euro-English and especially on lexical variations (e.g. Modiano, 2001, 2003; Simigné Feny, 2003; Murray, 2003; Mollin, 2006; Seidlhofer, 2004) has demonstrated that EU institutional communication arouses scientific interest and attention. Moreover, in 2016, senior translator at the European Court of Auditors, Jeremy Gardner, published an official report on the vocabulary developed by European institutions that often differs from that of any 'native' form of English. He included words that do not exist or are relatively unknown to native English speakers outside the EU institutions and words that are used with a meaning, often derived from other EU languages, that are not usually found in English dictionaries. Gardner (2016) reported the risk of a realistic miscommunication for both English native and non-native speakers: "the European institutions also need to communicate with the outside world and our documents need to be translated – both tasks that are not facilitated by the use of terminology that is unknown to native speakers and either does not appear in dictionaries or is shown in them with a different meaning. Finally, it is worth remembering that, whereas EU staff should be able to understand 'real' English, we cannot expect the general public to be *au fait* with the EU variety" (Gardner, 2016: 4).

This is also true for the continuous process of formation and reformulation of the terminological dimension of migration law metalanguage. The constant transformation of the semantics of specialized terminology due to the changing nature of migration processes and policies within the EU, the incomplete standardization of an official lexicon and the need for the interlingual intelligibility of this terminology system generate an urgent need to rethink and reinterpret ways and approaches to lexicographic resources able to represent the multifaceted nomination of European migration and asylum policy.

The intense political and regulatory activity which occurred as a result of issues arising from migration and asylum in each Member State was accompanied by a consequent proliferation of terminology. This led to the creation of multilingual glossaries, manuals and public, more or less official, lexicographic resources, which all aim to familiarise citizens, officers and migrants, with the lexical and conceptual aspects linked to the migratory context. In particular, the Directorate-General for Migration and Home Affairs of the European Commission has worked in recent years to encourage the comparison of terminology in many official languages of the European Union, as demonstrated by the compilation and publication of the Glossary on asylum and migration by the European Migration Network (EMN) which has been considered as a base for this study (Note 2). The Glossary is constantly updated, and the latest version (December 2022) includes approximately 540 terms and concepts in Euro-English reflecting the most recent EU policy on migration and asylum. It provides synonyms and translated terms into all EU Member States languages (English included) and a multi-word search tool.

3.2 Lexicography of migration and ELF corpora

What has been underlined so far relates directly to the linguacultural processes derived from the elimination of borders within the Schengen area, the proximity between cultures and the continuous exchange of people, knowledge and experiences over the past forty years. Therefore, issues and controversy related to legal as well as irregular migration across Europe have inevitably given rise to continuous political action in terms of policy, laws, and rules. As a consequence, growing and extensive work has been carried out in linguistics with respect to the words of migration, specifically focussing on the conceptual, semantic, and sociolinguistic aspects of their usage in different contexts.

Institutions around the world have developed databases aimed at defining the terms of greatest importance in the field of migration to allow consistency in their use and translation. The most important one in European contexts is the aforementioned EMN Glossary, which provides constantly updated information on the migration lexicon in EU institutions. By creating tools such as the EMN Glossary and the International Organisation of Migration's (IOM) Glossary on Migration (Note 3), political institutions attempt to foster and to promote the use of specialized terminology, which is expression of a specific political tendency. However, no indications are given about how the words presented in these glossaries might change when used in actual speech and daily exchanges in specialized contexts. In fact, tools like the official glossaries primarily focus on describing the words related to migration from a semantic dimension, to help the users understand the meaning of the words rather than their actual use in concrete communicative contexts such as international organisations, public administrations and services but also among individuals. In the Italian context, research studies such as Chiari (2021) and Longobardi & Pastorino (2023) testify a recent concern about lexicographic resources in migration discourse and their effective usability also in a translingual perspective.

As a matter of fact, authoritative multilingual resources such IATE (Interactive Terminology for Europe – Note 4) mainly focus on legal and administrative terminology related to EU regulations or best practices, often neglecting the deep linguacultural changes that intervene in the reception and interpretation of ESP terminology in each country. Moreover, most of the existing lexicographic resources are generally oriented toward EU languages and do not contain useful elements for interfacing and mediating with the languages of the beneficiaries of the legislation. This has direct consequences for the use of the available tools by mediators, officers, and operators. Issues and challenges raised by the migration lexicon are due to its internal stratification and to factors related to its production, i.e. international, national, regional, and local contexts, and to its reception, i.e. specialized (e.g. stakeholders, policymakers and operators) vs. non-specialized audience (e.g. migrants, refugees and asylum-seekers).

The complexity of the lexicographic dimension of migration discourse, especially in Europe, is inevitably related to the emerging role of ELF in national professional contexts where those terminological items are used and renegotiated by officers, migrants and asylum-seekers, and language mediators. Seen from this perspective, the substantial contribution corpus linguistics can give to lexicography seems crucial. As Firth (1957:11) stated with his famous recommendation: "you shall know a word by the company it keeps", the use and the context where a word is used should be consciously acquired before attempting to define and describe it.

So far, the study of global English language use, in general, and of ELF interactions, in particular, is supported by the availability of specific corpora which however are not specifically compiled for the purpose of representing ELF migration discourse. The International Corpus of English (ICE) (Note 5) provides representative samples of Inner and Outer-circle varieties of English with "the primary aim of collecting material for comparative studies of English worldwide" (Nelson, 2011). The 26 English varieties corpus (26EV) offers a comprehensive representation of written English language use around the world. On the other hand, the Asian Corpus of English (ACE) (Note 6), the Corpus of English as a Lingua Franca in Academic Settings (ELFA) (Note 7) and the Vienna-Oxford International Corpus of English (VOICE) (Note 8) are corpora of naturally occurring English used as a lingua franca by multilingual speakers, in a wide range of speech events: interviews, press conferences, service encounters, seminar discussions, working group discussions, workshop discussions; occurring in several domains such as education, leisure, professional business, professional organisation, and professional research/science. Even though these resources continuously provide data and material for identifying and describing characteristic language features of ELF and World Englishes, there is still an urgent need to develop lexicographic strategies and tools that cover significant gaps in the language of migration at both international and national levels, especially when English is involved.

4. Investigating linguacultural representations in specialized migration discourse from a lexicographic perspective

The investigation provided in this paper is only the preliminary step of ongoing and extensive work on the current lexicographic trends within the domain of migration discourse. Therefore, the data and results presented in section 5 are published with the intention of making available a preliminary methodological approach which can be used for further work both at the theoretical and analytical level. Starting from the assumptions derived from the previous theoretical background, the research objective is here to compare different lexicographic resources related to ESP migration discourse with a corpus of real and spontaneous ELF actualizations of specialized terminology, in order to inquire into the gaps and divergences in lexicographic tools, in terms of linguacultural representations and real uses of terms in professional contexts.

In order to reach this goal, first and foremost, the lexicon of migration has been considered according to three semantic macro-domains differentiated by specific pragmalinguistic features:

a) an international or transnational domain, identified with the institution of the EU rules and regulations. Terminology at this level is defined as specialized Euro-English since it needs to be standardized at least for all EU languages and to be consistent with the documentation found in specific EU national regulations (s. section 5.1);

b) a national or regional domain where procedures, practices, adaptations, and additions are expressed in the national language and proposed by each specific country administrative and general migration policies. Each EU country does in fact implement and define regulations and implementations in individual and not cross-nationally comparable ways. This level is affected by a certain degree of variability since migration policies, regulations and implementing decrees are constantly changing depending on government direction and overturns and on public opinion stances (e.g., in Italy, Security decrees have deeply changed in the last few years, according to the government policy). The second layer, being country-specific, represents the most controversial dimension in terms of usability, especially when ELF variations are used in communicative contexts where this terminology should be applied to real situations (s. section 5.2);

c) ELF-oriented local lexicon relevant to real and actual migration management processes concerning aspects that migrants, officers and mediators must face in their interactions with institutions and services such as social security, healthcare, education, administrative bodies (s. Section 5.3).

These three macro-categories determined the selection of three case-studies here presented by means of a contrastive analysis based on the comparison of different (lexicographic) tools, i.e. the *Oxford English Dictionary* (OED); the *International Organization for Migration* (IOM) *Glossary*; the *European Migration Network* (EMN) *Glossary*; the *NOW (News On the Web) Corpus*, a web-based collection of newspapers and magazines freely accessible online (Note 9); and a corpus of ELF and Italian Lingua-Franca variations adopted in spontaneous interactions among migrants, mediators and legal advisors in different Italian public centres for assistance and counselling to asylum-seekers and refugees, collected during a period of ethnographic research, between 2014 and 2019 (Sperti, 2017).

The items here presented have been analysed according to criteria of typology, frequency, semantic dimensions, lexicogrammatical creativity, and linguacultural representation. The choice of the different lexicographic resources and of the ELF authenticated expressions was made with the aim of investigating the gaps in the emerging ELF-oriented linguacultural representations in specialized migration terminology.

5. Linguacultural representations in specialized migration discourse: Euro-English vs. ELF-marked specialized terminology

Findings from analyses of various ELF corpora show that it is possible to advance the hypothesis that relatively constant dominant strategies of lexical variations emerge out of ELF speaker interactions in order to serve certain communicative functions. Pitzl et al. (2008), for example, reported different types of lexical innovations from VOICE. ELF speakers basically use these lexical innovations to increase explicitness and clarity, avoid potential ambiguity, reduce redundancy, regularize irregular forms and fill a lexical gap when speakers cannot recall a word.

What follows are the three case-studies under examination with the aim of comparing items pertaining to specialized terminology in migration both in Euro-English institutional contexts and in ELF-oriented local professional uses.

5.1 Euro-English and ELF authentication processes: residence permit vs. stay permit, permit of stay, permit to stay

The *residence permit* is one of the most used specialized terms in migration discourse since its role is crucial in the migratory process aimed to allow a migrant to stay legally in a host country. The OED's earliest evidence for *residence permit* is from 1846, in the writing of C. White. The term is simply defined as "permission to stay in a country".

In the IOM Glossary *residence permit* is "a document issued by the competent authorities of a State to a non-national confirming that he or she has the right to reside in the State concerned during the period of validity of the permit."

In the EMN Glossary *residence permit* is defined as "any authorisation issued by the authorities of a EU Member State allowing a third-country national to stay legally on its territory" (source: Regulation (EU) No 265/2010 (Long Stay Visa Regulation)). Synonyms listed are *residence document* and *single permit*, and related terms are *EU Blue Card* and *work permit*.

The *NOW Corpus* has been subsequently used to compare the Euro-English term *residence permit* with its ELF actualizations *stay permit, permit of stay, permit to stay*. Data show that *residence permit* is undoubtedly the most used in terms of frequency and distribution per country, but all the other three items are also reported, especially in Asian contexts (see Figures 1 and 2 below).

SECTION	ALL	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
FREQ	3340	29	69	64	93	81	100	209	262	252	401	431	370	485	494
WORDS (M)	18400	244.1	304.8	371.3	401.5	429.4	512.5	1,531.3	1,746.5	1,569.1	1,987.5	2,607.8	2,449.2	2,588.0	1,897.9
PER MIL	0.18	0.12	0.23	0.17	0.23	0.19	0.20	0.14	0.15	0.16	0.20	0.17	0.15	0.19	0.26

Figure 1. Frequencies of occurrence of *residence permit* in the *NOW Corpus*

In the ELF corpus from Italian migration contexts (Sperti, 2017) *residence permit* is never employed, while *stay permit, permit of stay, permit to stay* are interchangeably used in conversations, as shown in the following extract of an exchange:

(1) *every time that you come to the police to *renove it (.) they have to ask ehm to the commission (.) if it is possible to *renove the *permit to stay (.) ok?*

The ELF adaptation of *residence permit* is an example of what may happen to official terminology with an international scope when used in professional settings where users of English aim to mutual intelligibility by activating processes of linguistic creativity.

Frequency by country	(Return to frequency by year)			
SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
United States	350	7,269.5	0.05	
Canada	158	2,229.0	0.07	
Great Britain	517	2,574.2	0.20	
Ireland	222	1,256.1	0.18	
Australia	47	1,387.3	0.03	
New Zealand	54	680.3	0.08	
India	260	2,018.0	0.13	
Sri Lanka	22	142.8	0.15	
Pakistan	244	408.7	0.60	
Bangladesh	32	100.4	0.32	
Malaysia	71	398.2	0.18	
Singapore	133	642.7	0.21	
Philippines	41	515.0	0.08	
Hong Kong	88	90.9	0.97	
South Africa	423	832.6	0.51	
Nigeria	457	955.5	0.48	
Ghana	75	161.7	0.46	
Kenya	32	291.6	0.11	
Tanzania	28	38.9	0.72	
Jamaica	8	59.4	0.13	

Figure 2. Frequencies by country of *residence permit* in the *NOW Corpus*

5.2 Lexicogrammatical creativity in ELF spoken interactions: subsidiary protection vs. *sussidiarian protection

Dealing with international protection in migration discourse is often very challenging. This is particularly true in the EU context where rules in terms of protection are transposed in the national law of all Member States with a certain degree of variability.

As a result, in the OED the controversial term *subsidiary protection* is not listed and no reference to the specific meaning related to the legal dimension is mentioned for the adjective *subsidiary*. In fact, in the UK this kind of protection is generally called *humanitarian protection.* In the IOM Glossary the *subsidiary protection* is not included but it is specified that "at the regional level, the European Union uses the term "subsidiary protection" to refer to complementary protection granted to persons who are not covered by the Convention relating to the Status of Refugees ((adopted 28 July 1951, entered into force 22 April 1954) 189 UNTS 137) but are in need of international protection".

In the EMN Glossary *subsidiary protection* is defined as "the protection given to a third-country national or a stateless person who does not qualify as a refugee but in respect of whom substantial grounds have been shown for believing that the person concerned, if returned to their country of origin, or in the case of a stateless person to their country of former habitual residence, would face a real risk of suffering serious harm as defined in Art. 15 of Directive 2011/95/ EU (Recast Qualification Directive), and to whom Art. 17(1) and (2) of this Directive do not apply, and is unable or, owing to such risk, unwilling to avail themselves of the protection of that country" (source: art. 2(f) of Directive 2011/95/EC (Recast Qualification Directive)). Synonyms are *international protection* and *person eligible for subsidiary protection*, while related terms are *(civil) war refugee*, *humanitarian protection* and *temporary protection*.

In the *NOW Corpus* the term is quite rare and the occurrences are mostly retrieved in Irish sources; in Ireland this kind of protection is granted to asylum-seekers who are not recognized as refugees, but would suffer serious harm if they returned to their country of origin or of former habitual residence.

The ELF corpus reports an interesting example of regularization since *subsidiary* is changed to *sussidiarian (due to similarity to the Italian translation word *sussidiaria*) and easily used by all the participants involved in the exchange, as shown in the following extract:

(2) S1: *In Ghana you can come (.) with humanitarian *permit to stay and passport you can come back in your country (.) if you are a refugee or for *sussidiarian protection (..) in that case=*
S2: = But if you want to go to Africa they will give you problem they give you problem.

The extract shows that even those who are in charge of giving legal advice services adopt ELF lexical strategies, determined by their L1, to creatively fill a lexical gap without jeopardizing the conversational exchange which instead progresses successfully.

5.3 ESP terminology in migration discourse and ELF authentication processes: appeal vs. reappeal

The right to *appeal* is one of the most common reasons why an asylum-seeker asks for legal advice. The term, of Latin origin and borrowed from Old French, as reported by the OED, is one of the 5,000 most common words in modern written English. In the OED the verb *to appeal*, as regards the legal domain, is defined as "to call to a higher judge or tribunal for deliverance from the adverse decision of a lower; to remove a case formally from an inferior to a higher court".

In the IOM Glossary *appeal* is "a proceeding undertaken to have a decision reconsidered by a higher authority, especially the submission of a lower court's or agency's decision to a higher court for review and possible reversal" (source: B.A. Garner (ed), Black's Law Dictionary (10th edition, Westlaw, 2014)).

In the EMN Glossary the right to *appeal* is defined as "a general principle of European Union law set out in Art. 47 of the Charter of Fundamental Rights of the European Union which gives everyone whose rights and freedoms guaranteed by EU law are violated the right to an effective remedy before a national authority or tribunal notwithstanding that the violation has been committed by persons acting in an official capacity" (source: derived by EMN from Art. 47 of the Charter of Fundamental Rights of the European Union). A synonym is the *right to an effective remedy*, and related terms are *Charter of Fundamental Rights of the European Union* and *fundamental rights*.

Not surprisingly the *NOW Corpus* confirms the frequent use of the verb as well as of the noun *appeal* in different legal contexts worldwide (see Figures 3 and 4 below).

SECTION	ALL	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
FREQ	1310166	18207	23278	30056	35117	35377	42694	128647	139695	122827	164152	144875	122131	163736	139374
WORDS (M)	18400	244.1	304.8	371.3	401.5	429.4	512.5	1,531.3	1,746.5	1,569.1	1,987.5	2,607.8	2,449.2	2,588.0	1,897.9
PER MIL	71.20	74.57	76.38	80.95	87.47	82.38	83.31	84.01	79.98	78.28	82.59	55.55	49.87	63.26	73.44

Figure 3. Frequencies by country of *appeal* in the *NOW*

SECTION	FREQ	SIZE (M)	PER MIL	CLICK FOR CONTEXT (SEE ALL)
United States	246869	7,269.5	33.96	
Canada	117378	2,229.0	52.66	
Great Britain	177679	2,574.2	69.02	
Ireland	94983	1,256.1	75.62	
Australia	70387	1,387.3	50.74	
New Zealand	45924	680.3	67.50	
India	125151	2,018.0	62.02	
Sri Lanka	10067	142.8	70.50	
Pakistan	26286	408.7	64.31	
Bangladesh	6059	100.4	60.37	
Malaysia	44415	398.2	111.55	
Singapore	39411	642.7	61.32	
Philippines	23841	515.0	46.30	
Hong Kong	8045	90.9	88.48	
South Africa	73278	832.6	88.01	
Nigeria	146784	955.5	153.62	
Ghana	11395	161.7	70.47	
Kenya	27654	291.6	94.85	
Tanzania	2949	38.9	75.83	
Jamaica	5090	59.4	85.75	

Figure 4. Frequencies by country of *appeal* in the *NOW Corpus*

What is instead particularly interesting is the transformative process that may occur in ELF spoken interactions like the one that follows:

(3) S3: *my lawyer said I should *reappeal (.) he said I should *reappeal so that if I can reappeal they will give me back my document*
S4: *Mmm (.) but have you some paper about your *reappeal?*

The item *reappeal is constantly and successfully used throughout the corpus with the meaning of "appealing against the Territorial Commission's decision to reject the application for international protection at an Ordinary Court" by all the participants involved (migrants, legal advisors and mediators), regardless of their linguistic or educational origin. Once again, the creative process, activated by Italian ELF users, would suggest a L1 phonological transfer since the Italian translation of *appeal*, as meant in migratory discourse, is *ricorso*.

To sum up, the previous examples have confirmed that non-native speakers of English modify the language even at a specialized level. Therefore, the natural follow-up of this research should be a deeper investigation of the lexical innovations emerging from exchanges in migration contexts, both in spoken and in written forms, in order to define the exact role that these may have at the lexicographic level.

6. Conclusions

The present study has highlighted the influence that new linguistic landscapes and scenarios concerning the non-native use of English may have not only in everyday communication but also on the ESP terminology of migration as well as on its representation in lexicographic resources.

ELF data confirm that, thanks to international exchanges, speakers tend to use some existing ELF-based lexical items more frequently now than in the past; others appear to be undergoing a semantic redefinition and, new ones seem to have emerged, thus raising the issue of their representation in dictionaries and glossaries. Even if it is still too early to determine the real impact of ELF lexical variation such as to require a rethinking of lexicographic resources, the fact that ELF existing words and expressions are now being used successfully in cross-cultural interactions seems to suggest that a more extensive investigation including a wider range of ELF communicative dimensions is needed.

The analysis of the three case studies has shown different pragmalinguistic strategies applied to the construction of messages through ELF, rather than using standard varieties of English, in intercultural encounters involving migrants, mediators and legal advisors. Cogo (2010: 296) described ELF exchanges as those "where people from various backgrounds in more or less stable communities engage in communicative practices that shape, construct and define the communities themselves". Cogo claimed that lexicogrammatical norms are negotiated by speakers for specific purposes by providing a shared repertoire of resources that assure mutual understanding.

On the other hand, the previous discussion underlined the complexity of migration issues which deserve special attention by specialists in many fields of science and from different points of view, applied linguistics included. The rapid spread of English as a means of international communication, especially in migration contexts, requires a shift in theoretical, analytical, and pedagogical approaches to the study of English linguistics as well as of language learning, teaching and use. Lexicographic resources cannot be excluded from the current debate on lexical creativity and innovations in English. Detailed examinations of how speakers use the language as a lingua franca, in multilingual and ever-expanding interactional situations are more and more necessary. This is also true in the attempt to address the issues speakers with little or no competence in a foreign language face when trying to cope with the new linguistic and socio-cultural dimensions in the host country after leaving of fleeing from his/her home. At the same time, research studies in this field may help aiders and officers involved in migration contexts who could also benefit from lexicographic resources specifically prepared for institutional and interlingual communication.

It is unquestionable that ELF instantiations cannot be included in dictionaries or in other lexicographic resources, but the above discussion demonstrates that traditional terminology collections, such as glossaries or specialised dictionaries, are no longer able to meet and respond to their users' practical needs. A descriptive approach rather than a prescriptive one is required in the construction of lexicographic tools for migration discourse in the attempt to include all the semantic and the pragmalinguistic dimensions of a given term, especially when English is used as an international code to reach and connect people even in contexts where a specialized use of the language is required.

References

Baker, W. 2020. "English as a lingua franca and transcultural communication: rethinking competences and pedagogy for ELT", in C. Hall & R. Wicaksono (eds.), *Ontologies of English* (pp. 253-272). Cambridge: Cambridge University Press.

Baker, W. 2015. "Culture and complexity through English as a lingua franca: rethinking competences and pedagogy in ELT". *Journal of English as a Lingua Franca*, 4 (1), pp.9-30. De Gruyter.

Canagarajah, S. 2013. *Translingual Practice: Global Englishes and Cosmopolitan Relations*. Milton Park, Abingdon, Oxon; New York: Routledge.

Cheshire, J., Nortier, J. & Adger, D. 2015. "Emerging multiethnolects in Europe", *OPAL series*, Queen Mary University, London, #33.

Chiari, I. 2021. "A Lexicographic platform for migration terminology: problems and methods". *EURALEX Proceedings*, 665-672.

Cogo, A. 2010. "Strategic use and perceptions of English as a Lingua Franca". *Poznań Studies in Contemporary Linguistics*, 46,3: 295-312

Council of Europe 2020. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment. Companion Volume with new descriptors*. Strasbourg, Council of Europe.

Dendrinos, B. 2012. "Multi- and monolingualism in foreign language education in Europe", in Stickel, G. (ed.) *The Role of Language Education in Creating a Multilingual Europe*. Peter Lang.

Firth, A. 2009. "The lingua franca factor". *Intercultural Pragmatics*, 6: 147-170.

Firth, J. R. 1957. *Papers in Linguistics 1934–1951*. London: Oxford University Press.

Gardner, J. 2016. *Misused English words and expressions in EU publications*. European Court of Auditors.

Guido, M.G. 2018. "ELF in migration", in J. Jenkins, W. Baker & M. Dewey (eds.), *The Routledge Handbook of English as a Lingua Franca*, pp. 544-555. Routledge, London/New York.

Guido, M.G. 2015. "Mediating linguacultural asymmetries through ELF in unequal immigration encounters". *Lingue e Linguaggi*, 15, 155-175.

Guido, M.G. 2012. "ELF authentication and accommodation strategies in cross cultural immigration domains". *Journal of English as a Lingua Franca*, 1 (2), 219-240.

Guido, M.G. 2008. *English as a Lingua Franca in Cross-cultural Immigration Domains*. Peter Lang, Bern.

Hoffman C. 2000. "The spread of English and the growth of multilingualism with English in Europe", in Cenoz J. and Jessner U. (eds.), *English in Europe: The acquisition of a third language*, Multilingual Matters, Clevedon, pp. 1-2.

Hynninen, N. 2011. "The practice of 'mediation' in English as a lingua franca interaction". *Journal of Pragmatics*, 43, 965–977.

Jenkins, J., Cogo, A. & Dewey, M. 2011. Review of developments in research into English as a lingua franca, in "Language Teaching", 44, pp.281-315.

Jenkins, J., Modiano, M., & Seidlhofer, B. 2001. "Euro-English". *English Today*, 17(4(68)), 13 - 19.

Kramsch, C. 1993. *Context and culture in language teaching*. Oxford: Oxford University Press.

Longobardi, F. & Pastorino, V. 2023. "Labelled glossaries in the context of migration: a proposal for Italian and English". *De Europa*, 6(1), 61-84.

Lopriore, L. 2015. "What Competencies and Approaches for Multilingual Contexts? Mediation and Intercultural Communication in the Language Classroom". (219-230). In: Nied Curcio M., Katelhön P. Basic I. *Sprachmittlung/ Mediation/ Mediazione linguistica*. Berlin: Frank und Timm Verlagfur wissenschaftliche Lite.

Mauranen, A. 2018. "Conceptualising ELF". In J. Jenkins, C. Baker & M. Dewey (eds), *Routledge handbook of English as a lingua franca*, pp. 7–24. London: Routledge.

Mauranen, A. 2012. *Exploring ELF*. Cambridge: Cambridge University Press.

Modiano, M. 2003. "Euro-English: A Swedish perspective". *English Today*, 74: 35–41.

Modiano, M. 2001. "A new variety of English". *English Today*, 68: 13–16.

Mollin, S. 2006. "Euro-English. Assessing Variety Status". *Language in Performance*, 33. Tübingen: Narr.

Murray, H. 2003. "Swiss English teachers and Euro-English: Attitudes to a non-native variety". *Bulletin Suisse de Linguistique Applicée*, 77: 147–165.

Nelson, G. 2011. *The International Corpus of English*. http://ice-corpora.net/ice/.

Pitzl, M. L., A. Breiteneder & T. Klimpfinger (2008). "A world of words: Processes of lexical innovation in VOICE". *Vienna English Working Papers* 17,2:21-46.

Seidlhofer, B. 2011. *Understanding English as a lingua franca*. Oxford: Oxford University Press.

Seidlhofer, B. 2004. "Research perspectives on teaching English as a lingua franca". *Annual Review of Applied Linguistics*, 24: 209–239

Shohamy, E. and D. Gorter (eds.) 2009. *Linguistic landscape: Expanding the scenery*. New York: Routledge.

Simigné Feny, S. 2003. "The function of the English language in the European Union". *European Integration Studies*, 2: 53–64.

Sperti, S. 2017. *Phonopragmatic Dimensions of ELF in Specialized Immigration Contexts*. ESE: Salento University Publishing, Lecce.

Widdowson, H.G. 1979. *Explorations in applied linguistics*. Oxford: Oxford University Press.

Notes

Note 1. https://listen.politico.eu/EUConfidential189 (Feb 2021).

Note 2. https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary_en

Note 3. https://publications.iom.int/books/international-migration-law-ndeg34-glossary-migration

Note 4. https://iate.europa.eu/home

Note 5. https://www.ice-corpora.uzh.ch/en.html

Note 6. https://corpus.eduhk.hk/ace/

Note 7. https://www.helsinki.fi/en/researchgroups/english-as-a-lingua-franca-in-academic-settings/research/elfa-corpus

Note 8. https://voice.acdh.oeaw.ac.at/

Note 9. https://www.english-corpora.org/now/

Italian agri-food products abroad: word formation processes on Instagram

Sara Corrizzato[1]

[1] Department of Foreign Languages and Literatures, University of Verona, Verona, Italy.

Correspondence: Sara Corrizzato, Department of Foreign Languages and Literatures, University of Verona, Lungadige Porta Vittoria 41, 37139, Verona, Italy. E-mail: sara.corrizzato@univr.it

Abstract

The present study focuses on the identification and categorization of some cases of lexical adaptation and innovation, i.e. compounds, acronyms and borrowings, in the American Instagram accounts of nine agri-food Italian companies selling their products worldwide. The aim of the analysis is to investigate how and to what extent word formation processes in advertising discourse contribute to enhancing Italian products abroad and to promoting the 'made in Italy' concept. Examples of adapted and new terms gathered from the Instagram profile pages will be presented and examined within the promotional framework. Their presence will be then contextualised outside the American photo and video sharing social networking service to hypothesise their future lexicalisation.

Keywords: Instagram, Italian agri-food products, new word formation processes, promotional discourse

1. Introduction

The 'made in Italy' concept is often associated with high-quality craftsmanship, design, and style. Italy has a rich history of artistic and artisanal traditions, and many of its products are renowned for their quality and attention to detail (Fortis, 1998; Bucci, Coldeluppi & Ferraresi, 2011). The 'made in Italy' production is not limited to a specific industry but is commonly found in fashion, luxury goods, furniture, automotive, food and wine, and other sectors. Italian products are often valued for their creativity, innovation, and a combination of traditional craftsmanship. Overall, 'made in Italy' signifies a commitment to quality, authenticity, and a sense of style that has made Italian products popular worldwide (Bertoli & Resciniti, 2013a, 2013b).

During the second part of the 20th century, many Italian companies developed a growing interest in non-Italian markets, to the point that today they have a strong international presence, exporting their products around the world. The export of Italian goods not only promotes the country's economy but also enhances its cultural influence and brand recognition.

In the agri-food sector, Italy is known for its traditional food, which includes high-quality products. The Italian enogastronomic tradition is globally appreciated, as Italian wine, olive oil, pasta and cheese, among other products, are highly regarded. Hence, the 'made in Italy' label is also seen as a mark of authenticity and excellence (Beccattini, 2000, 2007).

For this reason, the promotion of agri-food products, as well as of other goods, has shifted from a national-only context to a much broader panorama, including product advertising to non-Italian audiences. This recent approach has led to the development of ad hoc strategies which means taking cultural differences and sensitivities into consideration, aligning with local tastes and preferences of the target audience and utilizing a variety of marketing channels that are popular and effective in the target markets. Parallel to this, the use of English has become of central importance to spread the advertising message in areas where Italian is not spoken by customers. (Bertoli & Resciniti, 2013a)

Considering that language can be seen as the primary tool to convey the promotional message, linguistic strategies in advertising are essential for effective communication, brand building and emotional engagement. Messages, which are supposed to be clear and concise, aim to help (prospective) customers understand the products, and choose them, creating a connection between the brand and the consumer. Therefore, by using terms and phrases that resonate with the target audience, advertisers can build messages which contribute to reinforcing the producer-customer relationship.

In the light of these considerations, investigating the way in which advertisers create the message can be important for two interconnected reasons. From a commercial perspective, linguistic strategies are essential for businesses to thrive in a competitive environment, attracting and retaining customers; from a more linguistic point of view, analysing the language used can be crucial to understand both how the message is moulded and, more specifically, how smaller units of meaning, such as words, may undergo changes to match the advertisers' creativity needs (Horninks & van Meurs, 2019).

In order to comprehend how specific words and expressions derived from advertisements can influence consumer perception, even to the point of entering the everyday lexical heritage, the next section aims to investigate how promotional message is built by the advertisers' skilful use of English. More specifically, the study focuses on the identification and categorization of new words, i.e. neologisms, acronyms and borrowings, which contribute to the successful promotion of Italian products worldwide. With this aim, the American Instagram accounts of nine agri-food Italian companies have been analysed: examples of new word formation gathered from the profile pages will be presented and examined within the promotional framework. Their presence will be then contextualised outside the American photo and video sharing social networking service to hypothesise their future lexicalisation.

2. Social media advertising and the role of language

As Charlesworth (2014: 66-67) clarifies, digital technology and social media have changed the relationship between companies and customers, allowing them to be closer and closer. Through the constant content sharing, customers have the opportunity to learn about the existence of products and companies can get in touch with them understanding their interests. Hence, social media have become a marketing tool of crucial importance. To promote their products, companies apply the principles of interactive marketing (Gretzel & Yoo, 2014), which opens up the possibility of building information, directly and collaboratively, with the customer, and therefore of providing customized solutions, by attracting and engaging (prospective) clients (Parson, Zeisser & Waitman, 1998; Hays, Page & Buhalis, 2012).

For this purpose, taking Instagram into account seemed a reasonable choice, as it ranks among the five largest social networks worldwide, with more than 2 billion users browsing the app each day (Note 1). Launched in 2010, Instagram is a word combination of 'instant' and 'telegram' and it was first started with proving functions to share and edit photos; later, many other functions have been added, such as videos, photo messaging, new filters and even the function of 'direct messaging', which allows users to exchange messages one-to-one privately. This content sharing platform applies a visual-based strategy and it is used by both private users and companies, as it allows the latter to get real-time metrics on how stories and promoted posts perform through the day and get insights into followers and how they react with the brand. Through Instagram brands can reach millions of online users and market their brand to tech-savvy audiences who look for shopping inspiration on their social media feeds.

Along with videos and pictures, language used in social media, and specifically on Instagram, plays a pivotal role both in shaping the message and in developing the producer-customer relationship, as it is the space in which linguistic creativity and non-standardization meet customers' attention (Milroy, 1991; Broderick & Pickton, 2005; Danesi, 2015, Epure & Mihăeş, 2018). Creating effective advertising content on Instagram, indeed, involves a combination of compelling visuals and engaging captions, with the ultimate aim at attracting the audience and making them remember the advertised product. The platform is designed to favour fast scrolling by the user so verbal messages have to be short and concise, choosing compelling language that resonates with the target audience as well as sharing the narrative behind the product to create a connection with them. As a consequence, words must be carefully selected to enable the customer to immediately recognise the product and the manufacturer.

In this regard, a considerable body of research (Loiacono, 2004; Kelly-Holmes, 2005; Jia-Ling & Bathia, 2008; Vettorel & Franceschi, 2019) agree on the fact that the language used in advertising "structures the vocabulary, [as well as] phraseological units" and "affects the speech characteristics of the individual" (Golubtsov et al., 2019: 117). From a linguistic perspective, this means that after a reasonable period of time, people tend to adopt the language used in their favourite promotional campaigns, replicating the same slogans in every-day life or using words which were earlier only associated with advertising. For this reason, identifying new or uncommon words and expressions coming from the world of advertising can help understand how language may change and be moulded.

3. Data and methodology

In order to carry out the analysis, the American Instagram pages of nine Italian agri-food companies were analysed. All the posts published by each company from December 2022 to December 2023 were taken into account, for a total of 1064 posts.

Table 1. Italian agri-food companies – number of Instagram captions

Italian agri-food company	American Instagram page	n. captions
Barilla (pasta, ready-made sauces and bakery products)	barillaus	226
Lavazza (coffee)	lavazzausa	196
Giovanni Rana (pasta and ready-made sauces)	giovanniranausa	72
Parmigiano Reggiano (dairy products)	parmigianoreggianousa	124
Ferrero Rocher (chocolate and confectionery products)	ferrerorocherusa	183
Allegrini (wine)	allegriniwinery	119

Kinder Ferrero (chocolate and confectionery products)	kinderus	141
Fratelli Beretta (cured meat)	fratelliberetta_usa	84
Mutti Pomodoro (ready-made sauces)	muttipomodorousa	19

Of each post, the verbal message included by the company just below the image was examined; neither the images nor any replies given by the company in response to user comments were examined. The selection was aimed to identify new words or expressions recurring in the posts examined.

The methodology chosen to categorise the selected material is adapted from Kuiper and Allan (1996). Clarifying that "the total vocabulary of a language is never static", Kuiper and Allan (1996: 181) offer an eye-opening classification to explain how new words enter the language and become part of it in a (more or less) permanent way, identifying three processes, i.e. compounding, affixation and conversion.

Kuiper and Allan (1996) also point out that lexical borrowing is one of the commonest ways through which a new word enters a language, i.e. when it is copied from the vocabulary of the original language into that of the target language. This phenomenon was common in the past when two communities of speakers came into contact, and it is even more common nowadays thanks to globalisation processes and the web. In addition to lexical borrowing, they include other two word formation processes, which result in blends and acronyms. The former are made by combining the first part of one word with the second part of another word, whereas the latter are created from the initial letters or syllables of a phrase.

As Crystal (1995) illustrates, coining new words is a natural linguistic process, but their survival cannot be taken for granted. A great number of coined words, indeed, are used temporarily by a small number of people and do not enter the language officially. If the coinage gains currency it means that the word has become lexicalised and it is used to conceptualise something new, as in the past when new words in the field of information technology were coined (Sari 2013) or, more recently, when now terms entered people's every-day communication during the coronavirus pandemic (Franceschi, 2021).

For the purpose of the present analysis, quantitative keyword research could certainly have speeded up the selection process through automated sampling, but given the nature of the texts, it seemed more appropriate to evaluate the corpus qualitatively. Besides being very short and concise messages, as explained above, the verbal component is always linked to the image or the video to which it refers. Disassociating the verbal element from the visual one could therefore have contributed to misinterpreting selected materials, not allowing the full comprehension of what was advertised by the companies. Moreover, most part of the examples presented in the next section only occasionally appear within the corpus. Therefore, the context was essential for the comprehension of the word.

In the next paragraph new words included in the American Instagram pages have been identified according to Kuiper and Allan's (1996) categorization and contextualised within their promotional framework. Parallel to linguistic categorization, their existence has been checked on the Merriam-Webster Dictionary and the Oxford English Dictionary.

4. Results

Before presenting the results from the qualitative analysis of the posts, it is worth making a brief premise, which could help better frame the type of advertising context which the Italian companies have chosen. In line with the discourse strategies used in the English version of Italian agri-food companies' websites (Corrizzato 2018, 2019), the typical 'made in Italy' imprint is preserved and strengthened throughout the publication of the posts, although promotion is not intended for the Italian public. Looking at the posts through a linguistic lens, this means that the language used in the verbal passages frequently evokes the idea of social gathering, emphasizing the importance of enjoying delicious food with family members. The advertisements suggest preparing an Italian meal on the occasion of typically American celebrations, for example the Halloween night or Thanksgiving Day (barillaus, parmigianoreggianousa, kinderus, lavazzausa), traditional celebrations, i.e. Christmas Eve (giovanniranausa, ferrerorocherusa, allegriniwinery) and Easter (kinderus), or more Italian festivals, such as the San Gennaro feast (muttipomodorousa). A special attention is also generally given to specific days of the year, in which the concept in the background is again to spend time with family members or close friends, as in the case of the National Linguine Day, the National Ice Cream Day (fratelliberetta_usa), the National Relaxation Day (barillaus), the National Coffee Day (lavazzausa) or the National Tree Week (muttipomodorousa).

In addition to that, references to the Italian origin of the ingredients, which is a guarantee of high quality, is frequently mentioned in the posts that promote Italian espresso (lavazzausa), Genovese pesto (barillaus) or pecorino Romano (giovanniranausa). The authenticity of the products is also guaranteed by the places that are mentioned, such as Piazza San Fedele in Milan (lavazzausa) and Allegrini's vineyards of the Oasi Mantellina (allegriniwinery).

Within this promotional framework, the qualitative analysis of the posts has revealed that the advertising strategies used to promote products, as well as their origin, draws on a series of morphological processes that make it possible

to include uncommon or new words in their narrative. Following Kuiper and Allan's categorization (1996) all the occurrences have been identified and classified, as the following paragraphs illustrate.

4.1 Borrowings

As Crystal (1995) maintains, borrowings are words taken from one language and incorporated into another. They play an important role in language evolution and development, as they express new concepts or indicate objects that may not have existed in the borrowing language. Besides linguistic enrichment, which is assured, borrowings also carry cultural connotations and nuances from the source language, providing speakers with a way to express concepts or practices that may not have direct equivalents in their own language. This is what happens in the Instagram posts which promote Italian products: in many cases they include Italian words to emphasise the quality and authenticity of the product.

From a lexicographical point of view, only a few words (ex. 1-3) have already become part of the American linguistic repertoire - *cavolo nero* and *tortelloni* are included in the Merriam-Webster Dictionary and *antipasto* in both the Merrian-Webster Dictionary and the Oxford English Dictionary. On the other hand, the other Italian terms (ex. 4-11) appear to be still unpopular, although they are massively used in the Italian cuisine. However, they cannot be found within dictionaries.

1. There's no better feeling than sharing a warm and cozy night in with friends—and no better recipe to do that with than this Pasta e Ceci recipe by @joechef_! With carrots, onions chickpeas, kale (or **cavolo nero**, for our friends across the pond!), warm vegetable stock, and plenty of delicious Ditalini, this is one winter soup recipe you won't soon forget. Get the full recipe below! (barillaus)

2. The gang's all here! Coppa, peppered salami, prosciutto- with their friends mozzarella, olives, crackers and more, waiting for you and your next get-together. The Fratelli Beretta Entertainment Line of **antipasto** platters have you covered! (fratelliberetta_usa)

Unlike the previous posts, the following caption presents a typical type of pasta from Northern Italy, *tortelloni*, which are similar in shape to *tortellini* but are usually larger in size and are generally filled with ricotta and spinach. Giovanniranausa offers a different version of *tortelloni*, using a very common adjective for the Italian audience, i.e. *asciutti*, which means that they are served with oil or melted butter, as explained in the post. (giovanniranausa)

3. This Christmas, our prosciutto & cheese **tortelloni** are once again the kings of the table. Adults and children, grandparents and grandchildren, tortelloni always please everyone with their classic and irresistible taste! The can be also prepared "**asciutti**" – i.e. without broth – then seasoned with melted butter and the company of a few fresh chervil leaves. (giovanniranausa)

Muttipomodorousa includes specific terms which pertain to the world of tomato processing: the four terms *pelati*, *concentrato*, *polpa* and *ciliegini* describe the four ways in which tomatoes are presented to customers.

4. Mutti's Peeled Tomatoes (**Pelati**) make this Salmorejo (Chilled Spanish Tomato Soup) a breeze! (muttipomodorousa)

5. Smooth, dense, and bold, our Tomato Paste (**Concentrato**) is the ultimate hack for adding the richness of Italian tomatoes to your dish. (muttipomodorousa)

6. For holiday cooking or everyday meals, our Crushed Tomatoes (**Polpa**) are sure to make dinnertime delizioso! (muttipomodorousa)

7. Mutti Cherry Tomatoes (**Ciliegini**) are the perfect addition to a delicious plate, bursting with sweet tomato flavor! (muttipomodorousa)

Interestingly, parmigianoreggianousa introduces a tool that is generally used to cut Parmisan cheese. Although in the Italian kitchen this type of knife is available, the clarification in brackets suggests that such a knife is not commonly used by non-Italian people (and this is confirmed by the fact that the term is not found in reference dictionaries).

8. Unkock the symphony of flavours! Have you noticed that funny little tool in our pictures, or maybe seen one being used in a restaurant? That's the "**coltello a mandorla**" (literally "almond-shaped knife") and it plays a crucial part in the sensory symphony that is Parmigiano Reggiano (parmigianoreggianousa)

The borrowing included in ex. 9 is a typical Italian expression and it is always used before lunch or dinner to wish the people who are eating with you a happy meal. Unlike *buon appetito*, which is traditionally rooted in the cultural context, the expression *aperi-cena* is a neologism that found its way into the Italian dictionary Treccani in 2012 to identify an eating moment which replaces dinner (Note 2).

9. @michelecasadeimassari is a curious culinary traveler – here is his quick and easy recipe for Chapati, an Indian flatbread that has become popular all over the world! **Buon appetito**! (parmigianoreggianousa)

10. Too hot to cook? :(Savor the summer vibes with a light and easy "**aperitivo**". All over Italy, it's traditional to have an assortment of light bites with a drink or two in the early evening… and sometimes it turns an "**aperi-cena**", a happy hour "snack-as-dinner!" (parmigianoreggianousa)

4.2 Affixation

As the selected examples show, affixation processes in this context turn out to be creative, requiring a completely intuitive understanding of them. In ex. 11 the two interjections *ooh* and *aah* are followed by the suffix -ing becoming two verbs which aim to symbolise the guests' astonishment.

11. #CookingInspiration that's easy to make, but will have all your dinner guests **ooh-ing and aah-ing**? Try there fancy Parmigiano Reggiano Sablés with Mortadella Mousee and Pistachios! (parmigianoreggianousa)

The same construction can be seen in ex. 12, in which the adjective *one* becomes *one-ish*, which is used to mean *approximately*. In this specific case, the linguistically creative solution is particularly relevant, as the *One-Pot Monday* is a key concept in the Barilla advertising campaign. Many of the culinary specialties promoted by the Italian agri-food company are in fact thought to be prepared cooking everything together at once. In the Barilla American Instagram page this hyphenated compound is always associated to Monday, which seems to be the day of the week when people have the least time to cook.

12. @chefrobthomas' recipe for #RastaPasta makes the most of seasonal grocery staples like peppers, onions, garlic, and of course, Barilla pasta! Let us know what you're cooking up for #OnePotMonday—or **One-ish Pot Monday** 😉—in the comments. (barillaus)

Seeking to understand the linguistic process that led to the construction of the name of the chocolates in ex. 13, it is possible to suppose that the marketing creators chose to preserve the noun *mandorla* (almond in English), which is the main ingredient of this confectionary product, and to add the suffix -ly, which produces the transformation of the noun into an adjective or adverb (i.e. *almondly*).

13. Ferrero Collection Grand Assortment is best enjoyed with loved ones. 😊 Which flavor are you most excited to enjoy together?
    […] **Manderly**… a delightful confection with a velvety hazelnut creamy filling and almond at its heart (ferrerorocherusa)

As shown in the ex. 14, creative affixation also works to combine two languages, English and Italian. The Italian inflectional suffix -o is added to the English root *classic* to intensify the connection of the place in which the company was born.

14. It's time to treat yourself to a new take on a **classic(o)** drink. This Café Mocha, paired with Kinder Bueno, is meant to be savored in a moment of unrushed bliss. (lavazzausa)

The last two posts of this category reverse the direction and show two examples where prefixes play a key role in building ad hoc words: in the first case the syllable *be* of the verb *believe* is replaced by the noun *brie* (ex. 15). Out of the culinary context the substitution would make no sense, but in this case, *brie* means soft cow's-milk cheese, which is one of the ingredients of the daily special.

The same linguistic transformation occurs in ex. 16, in which the adjective *beautiful* loses its first syllable to be replaced by the suffix *boo-*, evoking the spooky atmosphere of Halloween night.

15. For a Thanksgiving side that your dinner guests won't **brie-lieve**, try this #OnePotMonday Brie Mac & Cheese from @chelseapeachtree! Barilla Large Shells hold this delicious sauce for cheesy perfection in every bite. And for the topping, grated frozen cranberries! This tiny burst of color and flavor can turn a normal side dish into the star of the Thanksgiving spread. (barillaus)

16. Spooky craft time? Now that's what we call a **BOO-tiful** bonding moment! Let us know if you and your family will be making a sinisterly sweet paper chain craft or an at-home Boo Basket build! (kinderus)

4.3 Acronyms

The main purpose of acronyms within the corpus is to designate products that have been produced, processed and developed in a specific geographical area (e.g. the Italian acronyms DOP and DOC, and the English acronym DPO) or to indicate the quality of the product, such as EVOO olive oil.

In addition to that, barillaus in a few captions chooses acronyms to identify the main ingredients of a gastronomic delight. This is the case of PSL, which stands for Pumpkin (seed), (butternut) Squash, and (red) Lentil (pasta) or BTL sandwich, where B, T and L are the initial letters of its primary ingredients, i.e. bacon, lettuce, and tomato.

Although this language process is not widespread on the Instagram pages analysed, one acronym is worthy of examination, i.e. ASMR, since it is becoming very popular thanks to YouTube videos. The acronym, which stands for Autonomous Sensory Meridian Response, "is a sensory phenomenon, in which individuals experience a tingling, static-like sensation across the scalp, back of the neck and at times further areas in response to specific triggering audio and visual stimuli" (Barratt and Davis, 2015: 1). The increasing number of online video sharing platforms have contributed to facilitating the spread of this phenomenon, finding more and more people who

experience ASMR. As a consequence, a huge amount of videos have been produced and shared with the goal of provoking this pleasant sensation. As Barratt and Davis clarify, "ASMR videos also typically appear to include an emphasis on the use of sound to trigger the static sensation of ASMR, which include the subjects of these videos cycling through a variety of household items which make various noises when tapped upon or used" (2005: 2). Food-focused ASMR has also spread on the web, encouraging the diffusion of the phenomenon.

In this regard, parmigianoreggianousa enhances the qualities of Parmigiano Reggiano by focusing on the auditory characteristics associated with it.

17. Here's an **ASMR** classic for you – the sound of a rich, dense chunk of Parmigiano Reggiano being perfectly grated over some lucky person's bowl of pasta (parmigianoreggianousa)

18. Enjoy some "cheesy" **ASMR** this Saturday and listen to the bright, satisfying snare-drum sound of an expert inspector tapping a wheel of Parmigiano Reggiano for quality control (parmigianoreggianousa)

According to the Merriam-Webster Dictionary, the first known use of the word was in 2011, but a more in-depth analysis of the presence of the acronym in the food world would be useful to understand how frequently it is used and how the concept has adapted to the gastronomic context.

4.4 Compounds

Compounding allows languages to efficiently create new words by combining existing morphemes (Jurida, 2018). For this reason, in the language of advertising this linguistic process is widely used, as it allows content creators to work on new products, making them popular through innovative linguistic structures.

This is the case of the two compounds created by Barilla and Kinder Ferrero, although their uses on Instagram are different. *Pastastrology* (see ex. 19) was a recurring term on Barilla's Instagram profile during the year 2023, as the advertising campaign was based on the idea that each type of pasta can represent the personality of the person eating it. The open compound *Kinder tree* only appeared during the 2023 Christmas period, referring to the very popular song "Oh Christmas tree".

19. Just in time for the holidays, it's the ✨Barilla **Pastastrology**✨ Gift Guide with @itstaylorshapiro! Need some more gifting inspiration? Check out our Pastastrology personality alignment list below and share your pasta matches in the comments for a custom gift suggestion. (barillaus)

20. Oh **Kinder tree**, oh **Kinder tree**– how lovely are your Kinder Joy Holiday toy ornaments! Check out the link in our bio to create these magical holiday decorations. 🎄(kinderus)

The same attention should be given to the close compound *friendsgiving*, which officially entered the English language in 2007 (Note 3) but it is still not so popular. Used to indicate a meal shared with friends at Thanksgiving, the term frequently appears on the Instagram pages of the Italian companies, as examples 21 and 22 show:

21. Thanksgiving, **Friendsgiving**… whatever table you sit at, share the ravioli and tortellini recipes with the ones you love! (giovanniranausa)

22. We think that the best way to tell someone you love and appreciate them is to cook for them! And this Gratitude Season, we're making it easy with this Lemon Orzo & Kale Pasta Salad. It's perfect for game night, family holidays, or a **Friendsgiving** spread—and don't worry, there will be enough to go around. (barillaus)

4.5. Blends

The last category taken into consideration is blending. In example 23, the company Giovanni Rana from Verona builds a blend *astrorana* (astronaut + the Italian surname) to humorously refer to the United States Air Force.

23. **Astrorana.** What do astronauts eat when preparing for a space mission? The menu of Pastificio Rana, the star of a truly interstellar mission of course! To celebrate and honor Italian cuisine - a candidate for UNESCO's intangible heritage - the Rana chef team has developed an exclusive, super gourmet, and entirely Italian food program, which focuses on taste and nutrition: Ravioli stuffed with sweet Nerano lobster with courgettes fries and grated provola cheese (giovanniranausa).

Unlike the previous example, the blend included in the caption by Lavazza has gradually entered every-day communication, although it has not yet been included in dictionaries. *Coffeetail* (coffee + cocktail), in fact, is an umbrella term used by experts in the field to group cocktails in which the main ingredient is coffee. The analysis of the Lavazza Instagram page reveals a massive presence of the term, as coffeetail can be easily prepared at home.

24. Entertaining with #Lavazza 101: try a new **coffeetail** for the special holiday gathering. Brew fresh #LavazzaCoffee and mix up a variety of options for your guests, or craft a signature party drink. Which are you trying? (lavazzausa)

25. Have a martini glass, Lavazza Nitro Cold Brew, and the desire for a refreshing, not too sweet, chilled drink? This **coffeetail** will be ready in just moments – enjoy! (lavazzausa)

5. Conclusions

The linguistic analysis of the American Instagram profiles of Italian companies has led to the identification of two main trends. On the one hand, the promotional language, targeting non-Italian customers, aims to promote and reinforce the main concepts on which the 'made in Italy' label is built, i.e. the quality of Italian products, the authenticity of raw materials, the producers' know-how and emphasis on tradition.

On the other hand, the qualitative analysis of the Instagram captions has made it possible to recognise and classify all the new words or expressions identified. Every term examined in the present study belongs to the enogastronomic context (with the exception of the acronym ASMR) and can be divided into two categories: words created for a specific advertisement, and which are therefore expected to disappear in a short space of time, and words that are expected to have a long life in the linguistic-cultural context taken into account, as for *friendsgiving* and *coffeetail* or in the case of borrowings, e.g. *pelati*, *concetrato* e *aperi-cena*.

In this regard, this preliminary study has laid the groundwork for further research both from a lexico-semantic and lexicographic perspective, with the aim of understanding whether already existing or new words are likely to become a permanent part of the food lexicon. The drafting of a specialised glossary of innovative terms in this area could thus be the first step in an attempt to monitor the development of new formations.

References

Barratt E.L. & Davis N.J. (2015). Autonomous Sensory Meridian Response (ASMR): a flow-like mental state. *PeerJ 3*:e851, 1-17. https://doi.org/10.7717/peerj.851

Becattini, G. (2000). *Dal distretto industriale allo sviluppo locale. Svolgimento e difesa di un'idea*. Torino: Bollati Boringhieri.

Becattini, G. (2007). *Il calabrone Italia. Ricerche e ragionamenti sulla peculiarità economica italiana*. Bologna: Il Mulino.

Bertoli, G. & Resciniti R. (2013a). *International Marketing and the Country of Origin Effect: The Global Impact of 'Made in Italy'*. Cheltenham: Edward Elgar Publishing.

Bertoli, G. & Resciniti R. (2013b). Made in Italy e Country of Origin Effect. *Mercati e Competitività 2*, 13-36. http://doi:10.3280/mc2013-002002

Broderick, A. & Pickton D. (2005). *Integrated Marketing Communication*. Upper Saddle River: Prentice Hall.

Bucci A., Coldeluppi V., & Ferraresi M. (2011). *Il Made in Italy*. Roma: Carocci.

Charlesworth, A. (2014). *An Introduction to Social Media Marketing*. London: Routledge.

Corrizzato, S. (2018). Made in Italy provides the American Market with Added Value. A Linguistic Analysis of Wine Websites. *Iperstoria 11*, 145-155. https://doi.org/10.13136/2281-4582/2018.i11.547

Corrizzato, S. (2019). From the Local to the International Market: A Linguistic Analysis of Italian Agri-Food Products Promotional Discourse in the USA. In N. Bakić-Mirić et al. (Eds.), *Current Topics in Language and Literature - An International Perspective* (pp. 148-166). New Castle Upon Tyne: Cambridge Scholars Publishing.

Crystal. D. (1995). The Cambrisge Encyclopedia of the English Language. Cambridge: Cambridge University Press.

Danesi, M. (2015). Advertising Discourse. In A.F. Feed (Ed.), *The International Encyclopedia of Language and Social Interaction* (pp. 1-10). Hoboken: Wiley Online Library.

Epure, M. & Mihăeş L.C. (2018). *The Role of Language and Symbols in Promotional Strategies and Marketing Schemes*. Hershey: IGI Global.

Fortis, M. (1998). *Il Made in Italy*. Bologna: Il Mulino

Franceschi, D. (2021). Lexical-Semantic Representations at the Time of the Coronavirus Pandemic. International Journal of English Linguistics, 11(6), 120-129. https://doi.org/10.5539/ijel.v11n6p120

Golubtsov, S., Zalenskaya, V., Lominina, Z., Olomskaya, N., Uvarova, I., Bagova, I. (2019). The main methods of entering neologisms in modern advertising discourse. *Advances in Social Science, Education and Humanities Research 374*, 117-123. https://doi.org/10.2991/mplg-ia-19.2019.25

Gretzel, U., & Yoo K-H. (2013). Premises and Promises of Social Media Marketing in Tourism. In S. McCabe (Ed.). The Routledge Handbook of Tourism Marketing (pp. 491-504). London: Routledge.

Hays, S., Page, S.J., Buhalis, D. (2012). Social Media as a Destination Marketing Tool: Its Use by National Tourism Organisations. *Current Issues in Tourism 16*(3), 211–239. https://doi.org/10.1080/13683500.2012.662215

Horninks, J. & van Meurs F. (2019). *Foreign Languages in Advertising: Linguistic and Marketing Perspectives*. London: Palgrave Macmillan.

Jia-Ling, H. & Bhatia, T.K. (2008). Glocalization and English Mixing in Advertising in Taiwan: Its Discourse Domains, Linguistic Patterns, Cultural Constraints, Localized Creativity, and Socio-psychological Effects. *Journal of Creative Communications,* 3(2), 155-183. https://doi.org/10.1177/097325860800300203

Jurida, S.H. (2018). Word Formation in English: Derivation and Compounding. *DHS, 2*(5), 157-170.

Kelly-Holmes, H. (2005). *Advertising as multilingual communication*. New York: Springer.

Kuiper, K. & Allan W.S. (1996). *An Introduction to English Language: Word, Sound and Sentence*. London: Palgrave Macmillan.

Loiacono, A. (2004). *Linguistic Creativity in Advertising*. Bari: Uniongrafica Corcelli Editrice.

Milroy, L. (1991). *Language and Social Networks*. Hoboken: Wiley-Blackwell.

Parsons, A., Zeisser, M. & Waitman, R. (1998). Organizing Today for the Digital Marketing of Tomorrow. *Journal of Interactive Marketing 12*(1), 31-46. https://doi.org/10.1002/(SICI)1520-6653(199824)12:1<31::AID-DIR4>3.0.CO;2-X

Sari, R. (2013). Compounding In Information Technology Terms: A Morphological Study. *Bahasa & Sastra*, 13(1), 12-15. https://doi.org/10.17509/bs_jpbsp.v13i1.755

Vettorel P. & Franceschi V. (2019). English and other languages in Italian advertising. *World Englishes 38*, 417-434. https://doi.org/10.1111/weng.12432

Notes

Note 1. Retrieved December 30, 2023, from https://www.statista.com/topics/5286/instagram-marketing/#topicOverview

Note 2. Aperi-cena. In *Treccani*. Retrieved December 29, 2023, from https://www.treccani.it/vocabolario/apericena_res-2f149081-8991-11e8-a7cb-00271042e8d9_%28Neologismi%29/.

Note 3. Friendsgiving. In *Merriam-Webster Dictionary*. Retrieved December 29, 2023, from https://www.merriam-webster.com/dictionary/Friendsgiving.

Lexical Innovation in Ecotourism Discourse:

The Case of *Eco(-)lodge*

Lorenzo Buonvivere[1]

[1]Department of Foreign Languages, Literatures and Cultures, 'Roma Tre' University, Rome, Italy

Correspondence: Lorenzo Buonvivere, Department of Foreign Languages, Literatures and Cultures, 'Roma Tre' University, Via del Valco di S. Paolo 19, 00146 Rome, Italy. E-mail: lorenzo.buonvivere@uniroma3.it

Abstract

As repositories of the cultures whose language they describe, lexicographical resources partake in the (re)production of dominant ideologies. This is especially relevant with regard to the current ecological crisis. With this in mind, the present article contributes to research within the field of ecolexicography. Combining critical lexicography with ecolinguistics, it acknowledges the role of lexicographical resources in shaping the users' awareness of environmental protection. In particular, this study investigates lexical innovation within ecotourism discourse in order to understand whether "ecotourism talk" can respond to its sustainable objectives. The research focusses on one specific instance, the noun *eco(-)lodge*, which is examined by searching both native speakers' and learners' dictionaries and specialised and general English corpora. Results highlight a partial clash between the two types of sources. While examples of usage mostly connote ecolodges as a type of luxury and exclusive accommodation placed in natural – i.e., non-urban – contexts, dictionaries define them solely with reference to their supposed minimal environmental impact. Outcomes suggest a semantic bleaching of the combining form *eco-* in ecotourism discourse, which is exploited in lexical creations to advertise a form of niche tourism that does not always align with ecological concerns.

Keywords: ecolexicography, ecolinguistics, ecotourism discourse

1. Introduction

Ecotourism emerged as an opportunity to re-conceive travelling in times of unprecedented environmental breakdown. Included in the larger category of sustainable tourism, it refers to responsible travel to natural areas that tries to minimise impact on the destination's ecosystem, finance conservation projects, support local communities, and raise ecological awareness in its practitioners. Conservation, education, ethics, sustainability, and local benefits are the variables that are most frequently addressed in its definitions (Fennell, 2001). The former objectives may be imposed on a broad range of activities afforded to aspiring ecotourists, often in the form of adventurous experiences in natural environments, such as kayaking, cycling, hiking, safaris or other encounters with wildlife, etc. Despite the critics' concern with its factual sustainability and the persistence of (human) self-interest as opposed to the well-being of ecosystems, 'ecotourism is alive and well' (Fennell, 2022, p. 2), and actually growing (Precedence Research, 2023).

In terms of language, the specialisation of tourism discourse has long been ascertained (Gotti, 2006; Maci, 2020; Nigro, 2006), and studies have also explored the creativity of its lexical constructions (Frank, 2021; Koval et al., 2023) and suasive features (Manca, 2016) to arouse the interest of potential tourists. However, things ought to be reconsidered in the case of ecotourism communication, since its aim is not limited to the promotion of travel *per se* (or in order to maintain the host country's revenue flow), but it must encourage conscious behaviour in all actors involved, at all stages. Therefore, and since it has been developing its own 'register' (Dann, 1996a, 1996b), and producing more and more labels to describe the alternative experiences it offers (Rață et al., 2012), ecotourism discourse offers an interesting opportunity to address the ecolinguistic objective of 'explor[ing] the role of language in the life-sustaining interactions among humans, other species, and the physical environment' (Stibbe, 2021, p. 11). Indeed, given that interpretation, seen as the ability to understand and appreciate natural places and subjects, is one of ecotourism's key principles (TIES, 2019), the quality and performance of ecotourism activity is inextricably tied to the quality and performance of its discursive practices (Mühlhäusler, 2020; Mühlhäusler & Peace, 1999).

Therefore, the present paper investigates lexical innovation in ecotourism discourse in order to assess its function in supporting the ecological ends of the activity. It looks at the specific case illustrated by the noun *eco(-)lodge* (Note 1), created to advertise alternative types of accommodation in ecotourism destinations. The word features consistently in the official names of several structures, and ecolodges in general are seen as developing their own marketplace (Spenceley & Rylance, 2022). The analysis collocates itself within the theoretical framework of ecolexicography, which acknowledges the contribution that lexicographical practice and outputs can make to environmental protection. Section 2 provides a sketch of the former's history, although it devotes more attention to Liu et al.'s (2021) recent theorisation, as it guided the present research. Furthermore, it reviews some essential

literature contributing to the field. Section 3 proceeds to explain the methodology used, which relied on quantitative and qualitative approaches, and considered two main sources of data: dictionaries and corpora. Results are presented and discussed in Section 4. First, the meaning of *eco(-)lodge* is explored according to available definitions. Second, further insight into its context of use is obtained through the analysis of occurrences in the corpora. Findings from the two sources are compared to reveal a gap between them, offering a starting point for reasoning on common expectations about ecolodges. Whereas defined only with reference to their environmental impact, ecotourism texts mainly connote these buildings as luxury facilities found in "protected" natural contexts, generally overlooking sustainability. This leads to some considerations about the progressive semantic bleaching (Mühlhäusler & Peace, 2001) of the combining form *eco-* as a driver of lexical creativity. To conclude, section 5 summarises the outcomes of the study. It reflects on the viability of updating current lexicographical materials, and advocates for the need to redress the imbalance between economy and environmentalism evidenced in ecotourism discourse.

2. Theoretical Background

Ecolexicography was first proposed by Sarmento (2000, 2001, 2002, 2005), who intended to combine lexicographical research with the concerns of ecolinguistics (Fill & Mühlhäusler, 2001; Ha, 2023; Penz & Fill, 2022; Stibbe, 2021). His programme did not consist in the compilation of a specialised dictionary of ecological terms. Rather, he envisioned a project focussed on the identification, description, and categorisation of words according to their ecological or non-ecological role, that is, the favourable/unfavourable ideological stance they convey with respect to the preservation of the planet. Sarmento did not provide a clear criterion for distinguishing between the two classes of expressions, but simply drew up a list of examples to illustrate his point. For instance, *web of life*, *other animals*, and *in-oneself-with-the-other* were marked as having an ecological role; instead, *anthropocentrism*, *crisis of fragmentation*, and *Nature* were seen as non-ecological.

Sarmento set the agenda for the new (sub)discipline, even though factual contributions referring to his theory were not provided in the following years. The notion of ecolexicography was eventually taken up again by Albuquerque (2018a, 2018b, 2019a, 2019b). First, he discussed it as an instance of pedagogical lexicography, and conceived it as functional to the achievement of 'ecoliteracy' ('ecoletramento') (Albuquerque, 2018a). He postulated for the education towards ecological sensibility of both learners and teachers, starting from metalexicographic activities aimed at evaluating learners' dictionaries used in class. Later on, he recovered and developed Sarmento's original endeavour (Albuquerque, 2019a, 2019b). In fact, Albuquerque argued for the creation of an 'ecodictionary' ('ecovocabulário') which expands entries to address the consequences that using certain words may have on the relationships among individuals and ecosystems, so as to produce entries which foster an ecological worldview. The latter would consist of four main elements: definitions, examples of usage, effects ('efeitos') and results ('resultados').

More recently, Liu et al. (2021) put forward a new paradigm for doing ecolexicography, which is the one inspiring the present study. They distinguished between a micro and macro level of analysis. Micro-ecolexicography sees dictionaries as complex and 'integrated whole[s]' (p. 292) that can be compared to living ecologies. Hence, it seeks to investigate the interactions among their elements, both semantic (or strictly linguistic) – definition, examples, usage labels/notes, spelling, pronunciation, etc. – and semiotic (especially in e-dictionaries) – size, medium, interface, voice recordings, pictures, videos, hyperlinks, etc. Moreover, lexicographical resources participate in the "ecology" of distributed cognition, according to which learning results from the interplay among different materials and actors (including lexicographers). The prime objective of micro-ecolexicography is to account for all such variables in the design of effective dictionaries.

Conversely, macro-ecolexicography focusses on the relationship between dictionary compilation and ideology, from an ecological perspective. It considers the role of lexicographical resources in promoting environmental awareness, based on the users' response to available definitions. As such, it draws from the tradition of critical lexicography (Kachru & Kahane, 1995) and Critical Lexicographical Discourse Studies (Chen, 2019). These understand dictionaries as discourse, and accordingly recognise that they 'are not value-free representations' (Liu et al., 2021, p. 298) of the world, but rather 'constitute a repository of the common values and interests of the society whose language is described' (Béjoint, 2010, p. 202). Indeed, cultural data are reflected in dictionaries, whose entries disclose information as to the forms and meanings that the group of speakers represented there 'consider to be (in)appropriate, (un)desirable, (dis)agreeable, or (non)standard' (Łozowski, 2017, p. 166). Macro-ecolexicography is concerned with the ideological elements in dictionaries that reveal dominant stances about how to engage with the non-human world, including natural materials, non-human animals, plants, ecosystems, and so on. Therefore, it involves two types of complementary practices. One is metalexicographical, and consists in the investigation and critique of current entries to evaluate their (mis)alignment with worldviews based on ecological consciousness. The other is genuinely lexicographical, since it comprises the production or rectification of 'dictionary discourse to avoid biased representations of reality' (Liu et al., 2021, p. 305) (Note 2).

Within the history of ecolinguistic research, perhaps one of the first studies to account for the ecological role of words is Trampe's (1991/2001). In his work, Trampe reviewed a collection of lexical items retrieved from publications on agribusiness, and identified four main discursive tendencies in the domain: 1) reification, or the representation of living beings as objects or commodities, through nouns and verbs associated with manufactured

goods, such as *produce*, *manage*, or *recipient material* (to denote cows or sows to be implanted with an embryo); 2) 'euphemisation', or the use of euphemisms to conceal unpleasant facts about the industry, such as *plant protection tool* in place of *pesticide*; 3) denigration of traditional agriculture, described through such items as *rural*, *peasant-like*, or *retrograde*; 4) recurrence of slogans or phraseological elements that advertise intensive farming as inevitable and beneficial, e.g. *grow or drop out*, *ecology or economy* (Note 3).

Several other studies have followed Trampe's (1991) example, exploring recurrent lexical uses, either in environmental discourse or other domains, and critiquing the ideologies they imply or their effectiveness in raising environmental awareness. These include, for instance, the naming of the climate crisis (Liu & Huang, 2022; Penz, 2018), or the proliferation of compounds to talk about carbon production (Dury, 2008; Koteyko et al., 2010; Nerlich & Koteyko, 2009). However, because they do not consider lexicographical resources, they might be seen to contribute to what Sarmento (2001) called 'ecolexicology'.

Heuberger's (2003) is perhaps the study which best exemplifies the aims of ecolexicography. He investigated anthropocentrism in both native speakers' and learners' English dictionaries, searching for entries about non-human animals. He found that, in many cases, definitions presented them with exclusive reference to their usefulness, edibility, beauty, and harmfulness from a human perspective. On the contrary, more objective features such as appearance or typical behaviour were overlooked. For instance, an entry for *chicken* would read: 'a domestic fowl bred for its flesh or eggs, esp. young' (p. 92); or *vulture*: 'a large ugly bird with an almost featherless head and neck, which feeds on dead animals' (p. 97). Therefore, Heuberger (2003) concluded that dictionaries reflected the dominant view shared in most capitalistic societies, which sees non-human animals almost exclusively as resources to be exploited.

A similar trend was observed by Ezzati and Gholinejad Pirbazari (2022) with regard to plants. They reviewed all entries related to plants contained in the Persian dictionary *Farahng-e Bozorg-e Sokhan*. Their findings showed that, first, definitions often neglect careful taxonomy and categorise most plants vaguely as either "tree" or "bush"; second, they largely focus on characteristics that are appealing to humans in terms of edibility and/or therapeutic properties, resulting in a distinction between "useful" and "ornamental" plants. Moreover, entries consistently evaluate fruits and flowers by personifying them through adjectives such as "pretty", "ugly", "good-natured", etc. Taking earlier examples into consideration, the present study tries to expand on the former type of research and directly contribute to Liu et al.'s (2021) macro-ecoloxicography. Although it suggests some possible lines of intervention (see Section 5), it is primarily concerned with the critical stage of the ecolexicographical enterprise.

3. Methods

This research investigates the noun *eco(-)lodge* within ecotourism discourse. It aims to 1) explore its meaning according to available definitions and common uses; 2) evaluate its role in contributing to the promotion of responsible travel, and thus discuss the potential of lexical creativity in shaping ecological consciousness.

The methodology draws from a longstanding tradition of cooperation between lexicography and corpus linguistics (Faaß, 2018; Kilgariff, 2022; Rees, 2022). It combines quantitative and qualitative approaches in analysing and comparing data gathered from primary lexicographical resources, i.e., dictionaries, and English corpora.

To begin with, both native speakers' and learners' dictionaries were consulted. The former included the online editions of the *Oxford English Dictionary* (OED) and the *Merriam-Webster Dictionary* (MWD), in order to account for both the British and North-American contexts. As for the latter, all the online versions of the "Big Five" were searched – the *Oxford Advanced Learner's Dictionary* (OALD), the *Longman Dictionary of Contemporary English* (LDOCE), the *Collins Dictionary* (COBUILD), the *Cambridge Advanced Learner's English Dictionary* (CALD), and the *Macmillan English Dictionary for Advanced Learners* (MEDAL) (Note 4).

Following the retrieval of definitions provided by the abovementioned dictionaries, attention was turned to corpora to investigate examples of usage as found in actual texts. In light of the type of discourse evoked by the noun under scrutiny, the *EcoLexicon English Corpus* (EEC) (Faber et al., 2014) was initially targeted. The corpus was compiled by the LexiCon research group at the University of Granada. It contains contemporary environmental texts ranging between 1973 and 2016, and covering different domains such as biology, ecology, meteorology, environmental engineering, environmental law, etc. It totals 23.1 million words. After that, and with the aim of exploring the distribution and meaning of *eco(-)lodge* among non-specialised users, attention was turned to general corpora of English. Three main datasets were accordingly chosen: the *British National Corpus 2014* (BNC2014); the *Corpus of Contemporary American English* (COCA); and, finally, the *English Web Corpus 2021* (enTenTen21). In all corpora, all three forms of the lexeme – *ecolodge*, *eco-lodge*, and *eco lodge* – were searched.

As a further methodological step developed to ascertain the frequency with which buildings labelled as *eco(-)lodge* are qualified with reference to sustainability, the concordances produced by the corpora that showed matches for the search word (see Section 4) were accordingly and thoroughly analysed. In the case of the COCA corpus, it was possible to search through all of them manually. Instead, due to the larger quantity of occurrences found, and in order not to compromise the quality of the examination, the analysis of the enTenTen21 was restricted to a randomly obtained sample of 100 concordances for each form. This allowed to count how many concordances addressed ecological characteristics of the ecolodge, either with regard to the design and construction of the amenity itself or to its management. The following are instances of segments which were marked as befitting the former category:

Example (1)
The extraordinary Costa Rica **eco-lodge** was designed in harmony with the environment constructed from naturally fallen trees and recycled materials. [enTenTen21]
Meanwhile, the **ecolodge** is a linchpin in the preservation of a 25,000acre conservation area. [COCA]

Finally, a collocation analysis was performed to gain insight into the type of items which most often co-occur with *eco(-)lodge* and therefore participate in its definition. Due to the small number of occurrences in the COCA corpus (see Table 2), this was performed on the enTenTen21 only, using the *WordSketch* function featured in *SketchEngine*. Furthermore, as opposed to concordancing, the whole corpus was targeted in this case, and attention focussed on modifiers.

4. Results and discussion

4.1 Dictionaries

Some interesting findings emerged from the initial observation that most dictionaries do not include any entry for *eco(-)lodge* (see Table 1), starting from native speakers' sources. Nevertheless, whereas a search in the MWD produced no result at all, the OED records the hyphenated form *eco-lodge*. It does not register it as a headword, nor does it provide a definition, but simply lists it among the examples of sense 2.d for the voice *-eco-* as a combining form, which reads: "Denoting people and things relating to travel to and tourism in areas of ecological interest". It figures at the bottom of an inventory comprising also *eco-traveller*, *eco-adventure*, *ecotravel*, and *eco-resort*.

Table 1. Dictionaries containing *eco(-)lodge* as headword

Dictionary		Has an entry for *eco(-)lodge*
Native speakers'	OED	X
	MWD	X
	OALD	X
	LDOCE	X
Learners'	CALD	✓
	COBUILD	✓
	MEDAL	x

Learners' dictionaries, on the contrary, allowed to obtain formal definitions of the word, albeit only two out of the Big Five include it as an entry. Both the CALD and the COBUILD prefer the form *ecolodge*, defined as follows:
A place for people to stay on holiday that is designed not to harm the environment where it has been built. [CALD]
A tourist accommodation facility designed to have minimal impact on the environment, often constructed as part of an environmental project. [COBUILD]

As shown, the two definitions indicate minimal environmental impact as the primary, if not only, characteristic denoting this type of amenity. The COBUILD, in particular, suggests involvement with other, non-touristic activities, but retains the focus on ecological protection. The same may be inferred with regard to the OED. While the latter gives no details as to the kind of elements which qualify the structure as such, it does so indirectly. In fact, with reference to sense 2.d quoted previously, it states that "this activity [tourism to areas of ecological interest] is designed to be favourable to the environment". Therefore, there is a clear indication that the meaning overlaps with sense 2.e of *eco-*: "Denoting products, programmes, etc., which promote environmental conservation". Ultimately, according to strictly lexicographical resources, sustainability emerges as the most important aspect of an ecolodge.

4.2 Corpora

Figures resulting from the corpora align with the picture obtained from dictionaries, which frames *eco(-)lodge* as still in the process of being lexicalised (see Table 2).

Table 2. Occurrences of *eco(-)lodge* in the selected corpora

	ecolodge*		*eco-lodge**		*eco lodge**	
	Absolute freq	Relative freq (per million)	Absolute frequency	Relative freq (per million)	Absolute freq	Relative freq (per million)
EEC	-	-	-	-	-	-
BNC2014	-	-	-	-	-	-
COCA	28	0.03	19	0.02	8	0.01
enTenTen21	3,090	0.05	3,590	0.06	2,482	0.04

Based on relative frequency, little or no difference is detectable among the four corpora. In fact, in the present case, raw frequencies represent a more relevant value to take into consideration. Provided that the great imbalance observed is mainly due to the very different sizes of the corpora searched (EcoLexicon = 23.1 million words; BNC2014 = 112.1 million words; COCA = 1 billion words; enTenTen21 = 52 billion words), raw numbers still afford the opportunity to make some interesting considerations. The complete absence of *eco(-)lodge* in the *EcoLexicon English corpus* is particularly striking. Similarly, no occurrences of either form are found in the BNC2014, as the oldest and least recently updated of the four. The COCA, which collects texts up until 2019, does register some instances – albeit modest. On the contrary, the most conspicuous figures belong to the enTenTen21, comprising texts available online as of January 2022. Based on such a preliminary sketch, it may be speculated that *eco(-)lodge* is indeed a relatively recent creation, and that it mainly informs online communication as opposed to other media.

The second part of the analysis moved to the examination of concordances in order to assess how frequently discussions of ecolodges co-occur with mentions of their ecological impact. Table 3 below summarises the results obtained.

Table 3. Concordances mentioning sustainability/total concordances

	ecolodge*	*eco-lodge**	*eco lodge**
COCA	25/28	10/19	5/8
enTenTen21	35/100	36/100	22/100

At a first glance, the comparison reveals that while most of the occurrences in the COCA contain references to sustainability, they do not in the enTenTen21. In spite of this, results gained from the former must be interpreted in light of the representativeness afforded by the sample under consideration. The 28 hits obtained for *ecolodge**, for example, appear in 7 texts only. Out of the 25 instances in which ecological matters are addressed, as many as 18 are featured within the same article, titled "The eco home"; 5 belong to an environmental magazine; and 1 appears in a contribution to an academic journal discussing forest preservation. Thus, results are influenced by the fact that most of the COCA instances are documented in texts with an explicit environmental focus.

Instead, the enTenTen21 shows a considerable low frequency of mentions of ecological features in concordances for *eco(-)lodge*. Based on such preliminary findings, it may be argued that while sustainability is certainly accounted for in descriptions of such amenities, it is not as significant a factor as expected. Indeed, the substantial majority of the concordances abstains from referring to it. In addition, even though the enTenTen21 does not provide a topic tag for all the texts it includes, distribution still highlights that, as opposed to the COCA, here most of the texts recording the word belong to the domain of travel and tourism (see Table 4).

Table 4. Distribution of *eco(-)lodge* in the enTenTen21 according to topic

Topic	*ecolodge* Frequency	*eco-lodge* Frequency	*eco lodge* Frequency
None	2,982	3,231	2,305
Travel and tourism	206	295	152
Multi-topic	8	20	4
Politics and government	7	8	7
Home and family and children	4	2	-
Economy and finance and business	2	8	2
Beauty and fashion	2	3	-
Nature and environment	2	3	2
Education	2	1	-
Science	1	3	-
Culture and entertainment	1	6	-
Arts	1	1	-
Technology and IT	-	3	2
Hobbies	-	3	1
Cars and bikes	-	1	-
Games	-	1	-
Health	-	1	2
Sports	-	-	1

With the aim of gaining better understanding of the contextual items that typically co-occur with *eco(-)lodge*, a collocation analysis was carried out at this point. The list of modifiers obtained for each form is particularly revealing (see Table 5).

Table 5. List of modifiers collocating with *eco(-)lodge*

Ecolodge Modifier	Frequency	*eco-lodge* Modifier	Frequency	*eco lodge* Modifier	Frequency
Sina	7	rainforest	30	Sabanero	5
Rica	8	riverside	11	Preseli	6
rustic	18	family-run	7	Selva	3
rainforest	5	solar-powered	5	Rica	10
		rustic	28	off-grid	4
		secluded	16	rainforest	11
		boutique	24	beachfront	5
		beachfront	5	riverside	3
		luxury	93	safari	9
		luxurious	31	Venture	6
		five-star	7	luxury	39
		jungle	18	boutique	8
		upscale	9		
		expedition	5		
		award-winning	28		
		tranquil	6		
		charming	16		
		resort	9		

Some "noise" or interference is produced here by elements that are used in the proper names of facilities mentioned multiple times within the corpus. For instance, *Sina* and *Selva* are part of "Mt. Sina Ecolodge" and "Selva Verde eco lodge". Also, *Rica* features in "Costa Rica ecolodge/eco lodge", since the corpus contains a few texts that specifically review and advertise different accommodations in this destination. These items were disregarded in the analysis, as it was concerned with modifiers containing information about the characteristics of ecolodges.
The remaining items are indeed more salient. Three main semantic classes of words may be identified, qualifying the ecolodge according to its location, environmental impact, and level of comfort. The first group is composed of topographical modifiers, mostly nouns, including *rainforest* – which is the most consistent and characterises all three forms – *riverside*, *safari*, *beachfront*, and *jungle*.

Example (2)

> In Panama City, glass towers and boutique hotels dominate, though a drive just outside the city reveals a world of *rainforest* **ecolodges**, thatched beach huts, bird-watching resorts, and coffee haciendas.
> Your second base is a remote, *riverside* **eco-lodge** where you can bird-watch, hike, horseback ride, or explore the rainforest canopy with your local guide.

The second group comprises collocates that pertain to the domain of sustainability. It registers only two instances: *off-grid*, which denotes the absence of connections to public utilities or power services like electricity or gas; and *solar-powered*.

Example (3)

> The *off-grid* **eco lodge** has been designed to maximise the panoramic views and harness the power of the sun.
> A stay in this true *solar-powered* **eco-lodge** in the wilderness, about an hour from Brus Laguna, costs just $10 a night.

Finally, the remaining group of modifiers refers to the ranking of the amenity, or rather its degree of opulence. Apart from *rustic*, they all suggest high-end status. Both the nominal and adjectival forms *luxury* and *luxurious* are included (together, they collocate 124 times with *ecolodge*), followed by *boutique*, *upscale*, and *five-star*.

Example (5)

> Starting and ending in San José, the two of you will ford the scenic Río Pacuare (one of the best white-water rafting rivers in the world), explore lush jungle, and enjoy downtime together in a *luxury* **eco-lodge**.
> Your stay at this *luxurious* **eco-lodge** includes gourmet Ecuadorian meals, non-alcoholic drinks, and all activities and guided tours.

House in the Wild is a private *boutique* **eco lodge**, tucked away on a 1000 acre private estate, Naretoi, on the edge of the Maasai Mara.

Frégate features a luxurious *five-star* **eco-lodge** offering the optimum in comfort and amenities that has become a favourite hideaway for Hollywood stars, with deluxe villas right on the foreshore to ensure each has a million-dollar sea view.

Hence, the collocation analysis seems to indicate that, for the most part, ecolodges are described as being built in non-urban contexts, especially in natural areas conceived of as wild or untamed, such as forests or jungles. As a matter of fact, no occurrences of the noun *eco(-)lodge* denoting amenities built in urban contexts were found. Their ecological impact is rarely emphasised. In contrast, they are often presented as exclusive and expensive types of accommodation, catering to a restricted target of well-off tourists.

Furthermore, a closer look at concordances for these collocates, reveals that comfort and ecology are mostly kept separate, so that when the luxury aspect of ecolodges is addressed, references to sustainability are less frequent. When both features do appear, the latter are actually associated with other amenities located in the same area as the ecolodge, like in the following examples:

Example (6)

We stayed in luxury, [sic] eco-lodges and a house made entirely of mud with no electricity and walls open to the elements.

The striking Malewa Wildlife Lodge [...] has been transformed into a stunning luxury eco-lodge [...]. Only 1km from Malewa Wildlife Lodge, Kigio Wildlife Camp is a superb all-suite camp [...] constructed from sustainable pine-wood, local earth from the conservancy, traditional thatch provided by the surrounding community and canvas panels – there has been minimal usage of cement and steel and the environment has been carefully preserved during its construction.

Also, seen against the productivity of the category of opulence, the adjective *rustic* stands out in the collocates list, and opposes the main idea of retreat to luxurious ecolodges. The presence of both *rustic* and *luxurious* would suggest that the abodes have no standards in terms of ranking, and may offer either sumptuous and costly stays or humble and economical ones. Again, the qualitative examination of concordances for the former item as a modifier of *ecolodge* and *eco-lodge* allowed to clarify such apparent mismatch. The following example shows that *rustic* is not interpreted as "modest", "rough", "plain", or "simple":

Example (7)

Perfect for families, you'll also find on the property a giant tree to climb, gym equipment, a tennis and basketball court, plus an advertised heated pool that is often too cold to use, possibly due to the cooler climate. Onsite spa services are available. It may be noted that this *rustic* **eco-lodge** offers more in the way of amenities than most other available hotels in the area.

In fact, "rustic ecolodges" offer as many first-class services as luxury ones, and are equipped with the same kinds of amenities, such as gyms, basketball courts, heated pools, etc. Rather, the semantic value of rustic seems to concern the location of the ecolodge, and is constructed negatively in opposition with "urban". That is, the main meaning of *rustic* in the examples is actually "far from town", and therefore the word may be included in the group of modifiers referring to the positioning of such accommodations in natural areas.

This reinforces the idea according to which "true" nature can be experienced solely by getting away from urban settlements and into the wilderness surrounding them, i.e., their literal "environment". Such a tendency is observable in yet another modifier collocating with *eco(-)lodge*, namely, *secluded*. In point of fact, whereas ecolodges are built with the aim of granting closeness and direct access to nature, the former adjective qualifies the structures as separate – perhaps even shielded – from the places they are found in. Therefore, proximity to nature is mediated by means of deliberate confinement, precisely because the former is perceived as untamed, and hence potentially dangerous. A possible explanation for the insistence on the great comfort of ecolodges in these texts is that luxury is offered as an adequate and necessary compensation for the uncomfortableness of travelling to such distant, wild, and perilous destinations.

Example (8)

Enjoy the clean air, peace and tranquility at this *secluded* **eco-lodge**. You will lay in a hammock and listen to birds sing while being served refreshments.

Luna Lodge is a *secluded* **eco-lodge** set in the pristine and captivating Osa Peninsula, near Corcovado National Park in southwestern Costa Rica.

4.3 The 'Eco-' in 'Eco(-)lodge'

The analysis of dictionaries and corpora highlights a discrepancy between definitions and uses of *eco(-)lodge*. Both sources identify the structures as tourist facilities. However, on the one hand, lexicographical entries present them as having low environmental impact, and qualify ecolodges with exclusive reference to ecological consideration. On the other hand, instances retrieved from texts tend to overlook issues of sustainability, and mainly characterise the buildings as accommodations for five-star sojourns in natural areas.

One possible reason for such a shift may be ascribed to the polysemy of the adjective *ecological* itself, which has currently two main meanings. The first – and oldest, dating back to Ernst Haeckel's coinage of the term in German at the end of the nineteenth century – relates to biology, or rather ecology as a science. In this case, it indicates involvement with the interrelationships among organisms, and between organisms and their habitat. The second meaning is more recent – the OED's first entry dates back to 1969 – and it stands as a synonym of *environmental*, that is, concerned with the impact (and damage) of human activity on the natural environment.

Both senses are potentially carried by the combining form *eco-*, which has been particularly productive as a prefix in the creation of new words since the beginning of the present millennium (Benz, 2001). However, it would seem that the proliferation of "eco-words" corresponds to a gradual semantic opacification of the pre-modifier. Dury (2008) registered such phenomenon as an instance of Meyer and Mackintosh's (2001) 'de-terminologization', the process by which grammatical, semantic, and pragmatic shifts may occur when non-specialists borrow terms from a specialised domain that has become of interest to the general public. In the case of ecology, the recognition of the importance of ecosystemic relationships in the well-being of the natural environment has led non-scientists and laypeople in general to adopt *eco-* to denote products and attitudes mindful of ecological principles, regardless of their actual link with the discipline. According to Dury (2008), this is observable in the fact that "eco-terms" that have spread in general language are not equally used in specialised language.

The flexibility of *eco-* has made the former highly creative in ecotourism discourse (Rață et al., 2012). Not only has it provided the activity with its name, but it has also inspired a rich set of lexical innovations to describe its actors (e.g. *eco traveller*), activities (e.g. *eco trekking*), and destinations (e.g. *eco camp*). These uses do not pertain to the science of ecology. However, neither do they strictly convey the sense of "environmentally-friendly" observed in the combining form's de-terminologization. Indeed, the case of *eco(-)lodge* seems to testify to a further semantic shift in the meaning of *eco-*. As observed by the investigation of the corpora, and especially in light of the collocation analysis performed, the prefix does not necessarily suggest ecological concerns in the construction and maintenance of the facility, but rather refers to its location in non-urban environments (e.g. "rainforest ecolodge"). These findings confirm Mühlhäusler and Peace's (2001) early assumption about the 'severe semantic bleaching' of *eco-*, which '[r]ather than referring to functional interrelationships between the inhabitants of an ecology, [...] has come to mean something like "having to do with nature"' (p. 378). Such "general" connection with nature is ultimately only topographical. Provided that is the case, pre-modification through *eco-* would be tautological, since ecotourism, by definition, already implies (sustainable) activities that specifically take place in non-urban contexts.

Moreover, this movement from "environmental" to "in natural environments" is documented by the textual distribution of *eco(-)lodge* noted in the present study. Occurrences of the lemma in the COCA mostly contain explicit reference to sustainability, and these are found in environmental texts which seem to describe in great detail the ecological – i.e., environmental – features of the buildings under consideration. In contrast, occurrences listed in the enTenTen21, which rarely address ecological matters explicitly, belong to (eco)tourism discourse.

It is possible that ecotourism texts exploit the semantic ambiguity of *eco-* to advertise ecolodges and cater to conscious travellers. Because the combining form has by now taken the meaning of "opposing damage to the environment", and it is this latter sense that has prompted most lexical creations in recent years, readers may assume that the sustainability of the accommodation is inherent in the very use of the noun *eco(-)lodge*. In part, this could justify the low frequency of "green concordances" and collocates pertaining to the domain of environmentalism that was observed in the enTenTen21. However, previous studies about sustainable tourism found that sustainability tends to be consistently promoted in discourse, and related lexical items often register as keywords (Malavasi, 2017). Instead, the present analysis has revealed that, at least as far as ecolodges are concerned, comfort, luxury, and (secluded) proximity to natural areas are prioritised over environmental impact.

Benz (2001) discussed the semantic transformation of *eco-* from scientific to general use in terms of politicisation, considering how vocabulary was being re-shaped so as to meet the growing awareness about and campaigning for the state of the natural environment. Accordingly, he concluded his review of lexicographical sources by stating that '*eco(-)* plays a significant role in the lexicon of current English, proof that green discourse is an important register of current English, proof that there are increasingly diverse ways to speak and write of the environment and care for the earth' (Benz, 2001, p. 169) (Note 5). As opposed to this, present trends observed in the discursive construction of *eco(-)lodge* point to a reverse process, that is, a de-politicisation of *eco-*, whose "environmental" sense is being tempered to vaguely mean "found in natural, non-urban areas".

5. Conclusions

The analysis of *eco(-)lodge* has revealed that the newly-formed combination has not reached complete lexicalisation yet. To begin with, most of dictionaries do not include it as headword or, as is the case with the OED, only register it as an example. Second, no agreement is found as to the main orthographical form, as *eco*

lodge, *eco-lodge*, and *ecolodge* are all documented (with a slight preference for the former two variants), and hint at different degrees of lexicalisation. Finally, a partial semantic discrepancy between definitions and uses is observed. Dictionaries qualify ecolodges with exclusive reference to their environmental impact, whereas discursive instances often overlook the former and instead emphasise location and level of comfort.

Possibly, the gap between dictionary entries and corpora may be explained chronologically. Considering the differences that emerge in terms of textual distribution between the COCA and the enTenTen21, with the latter being more up-to-date, it is possible that dictionaries might have drawn their data from earlier occurrences of the noun. In other words, prior to becoming predominant in ecotourism advertising, *eco(-)lodge* might have been used chiefly in environmental discourse. A sign of this may be found in the OED, which includes the following example of usage for *eco-lodge*, dated 1991: 'By next year they plan to have an ecolodge and research station that will include walkways suspended in the canopy of the rainforest to be used by tourists and researchers'. This instance seems to indicate a double function for the ecolodge: it hosts both tourists and scientists engaged in doing research in the same area where the structure is built (similar to what the CALD entry implies, see Section 4). For this reason, it is likely that the term came to be restricted to the field of tourism somewhat later. However, a further and systematic diachronic analysis of *eco(-)lodge* is needed in order to test this hypothesis.

Still, all the dictionaries examined register a specialised use of the noun, denoting a type of tourist accommodation. As available entries bind the definition to the facility's environmental impact only, they do not acknowledge the meanings retrieved from the investigation of the corpora. On the one hand, they might be seen as misleading, since they induce users – learners, in particular – to assume that all ecolodges are sustainable "by definition". As a result, they might fail in uncovering promotional tourism texts (and their producers), which greenwash their activities by exploiting *eco(-)lodge* as a catchword. Within ecolexicography, in case of controversial entries, or entries which support an unfavourable representation of reality, especially in terms of ecological consciousness, Liu et al. (2021) recommend the need for direct intervention through the modification or update of the former. They suggest five main ways of 'reframing the dictionary discourse: warning, commenting, refining, questioning and neutralizing' (p. 305). In this case, lexicographers could intervene by adding modal items – e.g. "generally", or "in most cases", placed at the beginning of the definition – to allow for exceptions. They could also add further illustrative examples, given their 'potential [as] lexicographic resources of cultural data' (Łozowski, 2017, p. 166). Examples would avert such "neutral" sentences as, for instance, 'I'm currently building an ecolodge' [COBUILD], allowing for others of a more evaluative character – e.g. 'You can pay £3,000 to stay in a luxury ecolodge' [CALD].

On the other hand, definitions of *eco(-)lodge* might be considered as being acceptable, regardless of the factual sustainability of the structures addressed as such. Accordingly, the distance between lexical entry and use may be preserved so that the former represents a "standard", enlisting the properties a building must possess in order to be considered an ecolodge. This aligns with Rees's (2022) comment on the interactions between dictionary and corpus, as he observed that corpus-based and corpus-driven approaches to lexicography should face the fact that a dictionary is 'often regarded as a repository of "correct" language use' (p. 391).

In conclusion, what is more revealing about the present analysis, is that it addresses some (linguistic) shortcomings of the ecotourism industry. Admittedly, comfort and sustainability are not mutually exclusive, and an ecolodge could be built to be both comfortable and sustainable. As a consequence, promotional materials will advertise environmental impact *and* other factors, such as ranking of the structure, services provided, and so on. Stamou and Paraskevopoulos (2003, 2004) understand ecotourism discourse as instantiating itself along a continuum which has tourism (i.e., economy) and environmentalism at its ends, so that elements of both discourses will always inform ecotourism texts. Results from the current research, however, register an imbalance between the two. In fact, discursive constructions emphasise luxury over sustainability, to the extent that the *eco-* in *eco(-)lodge* loses its environmental connotation and refers only to the location of the amenity.

Ultimately, the present study demonstrates how ecolexicography – intended not only as the production of lexicographical materials about ecological vocabulary, but rather as the critical-discursive examination of texts which include lexicographical resources – can be used to investigate the interrelationship between language and ecology, and assess the viability of its outcomes, particularly in terms of lexical innovation.

References

Albuquerque, D. (2018a). As relações entre ecolexicografia e lexicografia pedagógica. *Domínios de Lingu@gem*, *12*(4), 2066–2101. https://doi.org/10.14393/DL36-v12n4a2018-7

Albuquerque, D. (2018b). Revisitando a ecolexicografia. *Revista de Letras*, *37*(2), 135–147.

Albuquerque, D. (2019a). Novas perspectivas nos estudos lexicográficos: A ecolexicografia e as palavras ecológicas. *Revista (Entre Parênteses)*, *8*(1). https://doi.org/10.32988/rep.v1n8.845

Albuquerque, D. (2019b). Subsídios para um projeto ecolexicográfico: As palavras ecológicas e o ecovocabulário. *Afluente*, *4*(12), 42–63.

Béjoint, H. (2010). *The lexicography of English*. Oxford University Press.

Benz, B. (2001). "Eco" and the new millennium: Current coverage of "eco-" in dictionaries. *Dictionaries: Journal of the Dictionary Society of North America*, *22*(1), 163–170.

Chen, W. (2019). Towards a discourse approach to critical lexicography. *International Journal of Lexicography*, *32*(3), 362–388. https://doi.org/10.1093/ijl/ecz003

Dann, G. (1996a). Greenspeak: An analysis of the language of eco-tourism. *Progress in Tourism and Hospitality Research*, *2*, 247–259.

Dann, G. (1996b). *The language of tourism: A sociolinguistic perspective*. CAB International.

Dury, P. (2008). The rise of "carbon neutral" and "compensation carbone": A diachronic investigation into the migration of vocabulary from the language of ecology to newspaper language and vice versa. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, *14*(2), 230–248. https://doi.org/10.1075/term.14.2.06dur

Ezzati, E., & Gholinejad Pirbazari, B. (2022). An eco-linguistic study on plants: A case study on plant entries in Farhang-e Bozorg-e Sokhan. *International Journal of Agricultural Management and Development*, *12*(2), 159–168.

Faaß, G. (2018). Lexicography and corpus linguistics. In P. A. Fuertes Olivera (Ed.), *The Routledge handbook of lexicography* (pp. 123–137). Routledge.

Faber, P., León-Araúz, P., & Reimerink, A. (2014). Representing environmental knowledge in EcoLexicon. In E. Bárcena, T. Read, & J. Arús (Eds.), *Languages for Specific Purposes in the Digital Era* (pp. 267–301). Springer International Publishing. https://doi.org/10.1007/978-3-319-02222-2_13

Fennell, D. A. (2001). A content analysis of ecotourism definitions. *Current Issues in Tourism*, *4*(5), 403–421. https://doi.org/10.1080/13683500108667896

Fennell, D. A. (2022). Introduction. In D. A. Fennell (Ed.), *Routledge handbook of ecotourism* (pp. 1–6). Routledge.

Fill, A., & Mühlhäusler, P. (Eds.). (2001). *The ecolinguistics reader: Language, ecology and environment*. Continuum.

Frank, T. O. (2021). Neonym analysis in the English language for tourism purposes. *Journal of Teaching English for Specific and Academic Purposes*, *8*(3), 187–200. https://doi.org/10.22190/JTESAP2003187O

Gotti, M. (2006). The language of tourism as specialized discourse. In *Translating Tourism: Linguistic/Cultural Representations* (pp. 15–34). Università degli Studi di Trento: Dipartimento di Studi Letterari, Linguistici e Filologici.

Ha, C. (2023). The year's work in ecolinguistics 2022. *Journal of World Languages*, *9*(2), 231–252. https://doi.org/10.1515/jwl-2023-0001

Heuberger, R. (2003). Anthropocentrism in monolingual English dictionaries: An ecolinguistic approach to the lexicographic treatment of faunal terminology. *AAA: Arbeiten Aus Anglistik Und Amerikanistik*, *28*(1), 93–105.

Kachru, B. B., & Kahane, H. (Eds.). (1995). *Cultures, ideologies, and the dictionary: Studies in honor of Ladislav Zgusta*. Max Niemeyer Verlag. https://doi.org/10.1515/9783110957075

Kilgariff, A. (2022). Using corpora as data sources for dictionaries. In H. Jackson (Ed.), *The Bloomsbury handbook of lexicography* (2nd ed., pp. 71–88). Bloomsbury Academic.

Koteyko, N., Thelwall, M., & Nerlich, B. (2010). From carbon markets to carbon morality: Creative compounds as framing devices in online discourses on climate change mitigation. *Science Communication*, *32*(1), 25–54. https://doi.org/10.1177/1075547009340421

Koval, N., Kushka, B., Nagachevska, O., Valeriia Mykhaylivna, S., & Uhryn, L. (2023). Changing nature of English tourism discourse: A linguistic approach. *Arab World English Journal*, 215–225. https://doi.org/10.2139/ssrn.4350369

Liu, M., & Huang, J. (2022). "Climate change" vs. "global warming": A corpus-assisted discourse analysis of two popular terms in *The New York Times*. *Journal of World Languages*, *8*(1), 34–55. https://doi.org/10.1515/jwl-2022-0004

Liu, X., Lyu, J., & Zheng, D. (2021). For a better dictionary: Revisiting ecolexicography as a new paradigm. *Lexikos*, *31*(1), 283–321. https://doi.org/10.5788/31-1-1648

Łozowski, P. (2017). Dictionaries and culture. In P. A. Fuertes Olivera (Ed.), *The Routledge handbook of lexicography* (pp. 166–178). Routledge.

Maci, S. (2020). *English tourism discourse: Insights into the professional, promotional and digital language of tourism*. Editore Ulrico Hoepli.

Malavasi, D. (2017). 'No one can be the invisible tourist—But we like that you're trying': An analysis of the language of sustainable tourism. In M. Gotti, S. Maci, & M. Sala (Eds.), *Ways of Seeing, Ways of Being: Representing the Voices of Tourism* (pp. 363–377). Peter Lang.

Manca, E. (2016). *Persuasion in tourism discourse: Methodologies and models*. Cambridge Scholars Publishing.

Mattiello, E. (2022). *Transitional morphology: Combining forms in modern English*. Cambridge University Press. https://doi.org/10.1017/9781009168274

Meyer, I., & Mackintosh, K. (2001). When terms move into our everyday lives: An overview of de-terminologization. *Terminology*, *6*(1), 111–138. https://doi.org/10.1075/term.6.1.07mey

Mühlhäusler, P. (2020). Quo vadis ecolinguistics? *ECO-REBEL: Revista Brasileira de Ecologia e Linguagem*, *6*(1), 5–23.

Mühlhäusler, P., & Peace, A. (2001). Discourses of ecotourism: The case of Fraser Island, Queensland. *Language & Communication*, *21*, 359–380. https://doi.org/10.1016/S0271-5309(01)00006-4

Mühlhäusler, P., & Peace, A. (1999). Mind your language: Ecolinguistics as a resource for ecotourism. *Proceedings of the Ecotourism Association of Australia 1999*. National Conference of the Australian Ecotourism Association.

Nerlich, B., & Koteyko, N. (2009). Compounds, creativity and complexity in climate change communication: The case of "carbon indulgences". *Global Environmental Change*, *19*(3), 345–353. https://doi.org/10.1016/j.gloenvcha.2009.03.001

Nigro, M. G. (2006). *Il linguaggio specialistico del turismo: Aspetti storici, teorici e traduttivi*. Aracne.

Penz, H. (2018). "Global warming" or "climate change"? In A. F. Fill, & H. Penz (Eds.), *The Routledge handbook of ecolinguistics* (pp. 277–292). Routledge.

Penz, H., & Fill, A. (2022). Ecolinguistics: History, today, and tomorrow. *Journal of World Languages*, *8*(2), 232–253. https://doi.org/10.1515/jwl-2022-0008

Precedence Research. (2023). *Ecotourism market size to attain USD 656.19 billion by 2032*. https://www.precedenceresearch.com/ecotourism-market

Rață, G., Petroman, C., Petroman, I., & Percović, A. (2012). The English of Indian eco-tourism. In G. Rață (Ed.), *The English of tourism* (pp. 59–62). Cambridge Scholars Publishing.

Rees, G. (2022). Using corpora to write dictionaries. In A. O'Keeffe & M. J. McCarthy, *The Routledge handbook of corpus linguistics* (2nd ed., pp. 387–404). Routledge. https://doi.org/10.4324/9780367076399

Sarmento, M. S. (2000). Ecolexicography: Words and expressions we should live by. *Österreichische Linguistiktagung Graz, 8.–10. Dezember 2000. Akten Des Symposiums 30 Jahre Sprache Und Ökologie — Errungenschaften, Visionen*. Symposiums 30 Jahre Sprache und Ökologie — Errungenschaften, Visionen, Graz.

Sarmento, M. S. (2001). Ecolexicography: Ecological and non-ecological words and expressions. *Langscape Newsletter*, 20, 37–44.

Sarmento, M. S. (2002). Ecolexicography: Ecological and unecological words and expressions. In A. Fill, H. Penz, & W. Trampe (Eds.), *Colourful Green Ideas: Papers from the Conference 30 Years of Language and Ecology (Graz, 2000) and the Symposium Sprache und Ökologie (Passau, 2001)* (pp. 487–492). Peter Lang.

Sarmento, M. S. (2005). Por uma ecolexicografia. *Confluências*, *2*, 84–97.

Spenceley, A., & Rylance, A. (2022). Ecotourism and the Sustainable Development Goals. In D. A. Fennell (Ed.), *Routledge handbook of ecotourism* (pp. 9–22). Routledge.

Stamou, A. G., & Paraskevopoulos, S. (2003). Ecotourism experiences in visitors' books of a Greek reserve: A critical discourse analysis perspective. *Sociologia Ruralis*, *43*(1), 34–55. https://doi.org/10.1111/1467-9523.00228

Stamou, A. G., & Paraskevopoulos, S. (2004). Images of nature by tourism and environmentalist discourses in visitors books: A critical discourse analysis of ecotourism. *Discourse & Society*, *15*(1), 105–129. https://doi.org/10.1177/0957926504038948

Stibbe, A. (2021). *Ecolinguistics: Language, ecology and the stories we live by* (Second edition). Routledge.

The International Ecotourism Society [TIES]. (2019). *What Is Ecotourism?* https://ecotourism.org/what-is-ecotourism/

Trampe, W. (2001). Language and ecological crisis: Extracts from a dictionary of industrial agriculture. In A. Fill & P. Mühlhäusler (Eds.), *The ecolinguistics reader: Language, ecology, and environment* (pp. 232–240). Continuum. (Original work published 1991)

Notes

Note 1. In the present article, *eco(-)lodge* (in italics) refers to the lexeme searched for in both dictionaries and corpora, and thus includes all the possible forms in which the noun is documented: *eco-lodge*, *ecolodge*, and *eco lodge*; "ecolodge" (in normal typeset) is used in the discussion to indicate the type of building denoted generally by the former. Although mistakeable for a compound, according to transitional morphology (Mattiello, 2022), *eco(-)lodge* is an instance of combining form (CF) combination, obtained from the neoclassical CF *eco-* (of Greek origin). As explained by Mattiello (2022), CFs combinations may share some properties with compounds, such as compositionality, that is, the ability to be understandable from the meanings of their constituent morphemes. *Eco(-)lodge* belongs to the latter case.

Note 2. Because of the relationship between dominant discourses and lexical uses, Liu et al. (2021) include the compilation of dictionaries of endangered languages among the concerns of macro-ecolexicography.

Note 3. Liu et al. (2021) suggest that good ecolexicographical practice would ensure that dictionary entries avoid or rectify the use of reifications and euphemisms as observed by Trampe (1991).

Note 4. The online edition of the Macmillan Dictionary was officially taken down in July 2023. However, it was still available at the time when data for the present study were gathered.

Note 5. As a sign of the growing influence of environmental concerns in public discussions, Benz (2001) noted the gradual transformation of *eco-* from a bounded combining form to a free-standing adjective or even noun. The OED indeed registers the headword *eco* as an adjective with the meaning of 'environmentally friendly' (e.g. 'It seems logical for spas in the wild to be eco').