Uniwersytet Mikołaja Kopernika w Toruniu
Wydział Humanistyczny

Dorota Watkowska

A Cognitive (Construction) Grammar approach to written academic English as a lingua franca: the emerging elaborated abstract construction in three disciplinary communities

Pisany akademicki język angielski jako lingua franca w perspektywie kognitywnej gramatyki (konstrukcji): wyłaniająca się rozwinięta konstrukcja abstraktu w trzech społecznościach dyscyplinowych

Praca doktorska napisana pod kierunkiem
dr hab. Ariadny Strugielskiej, prof. UMK

Toruń 2024

Table of contents

List of tables

List of figures

After all this time? Always.
Dziękuję, Pani Profesor.

Introduction

The purpose of the present dissertation is to identify and explore the emerging elaborated abstract construction in three academic disciplinary communities, i.e. law, linguistics, and literary studies. The rationale behind formulating the above aim stemmed from current studies in two research fields, i.e. academic English as a lingua franca and Cognitive (Construction) Grammar, the integration of which was perceived as facilitating further advancements in both areas. To provide a detailed account of the avenues along which these developments could proceed, culminating in the emergence of the elaborated abstract construction, i.e. to establish a well-grounded motivation for the current project, the dissertation is divided into three main parts, each consisting of two chapters.

The first part is dedicated to (academic) English as a lingua franca. To be more specific, the first chapter focuses on defining the concept of a lingua franca and presenting the main theoretical approaches used in theorizing English as a lingua franca (henceforth ELF). The chapter commences with a description of lingua francas and ELF from a sociolinguistic angle, utilizing three continua of features employed to characterize lingua francas. In this way, ELF is presented as related to, yet still different from, other lingua francas, and its uniquely global role is duly emphasized. Next, the chapter focuses on discussing two approaches embraced in ELF theorizing, i.e. competence-based and usage-based, presenting complementary views upon the relationship between linguistic knowledge and language use. The first chapter concludes with an indication that a usage-based approach, with its emergent and emerging perspectives, appears most promising in ELF theorizing as it adequately captures an interplay of fluidity and stability inherent in ELF, particularly in the academic domain.

Consequently, the second chapter concentrates on describing academic ELF from a usage-based perspective. In this chapter, the role of ELF among academic (disciplinary) communities is discussed, highlighting ELF's crucial function in international research dissemination. Simultaneously, the usage-based perspective outlined in Chapter 1 is further elaborated through two models employed for theorizing academic ELF, i.e. Linear Unit Grammar (henceforth LUG) and Construction Grammar (henceforth CxG), corresponding to a more emergent and a more emerging perspective, respectively.

Subsequently, specific methodological solutions adopted for exploring academic ELF are delineated in two steps. Firstly, principles of constructing spoken and written academic ELF corpora are provided. Secondly, methods of analyzing spoken and written data in accordance with LUG and CxG are discussed. The chapter ends with a suggestion that a more comprehensive account of academic ELF can be achieved through integrating the principles of LUG and CxG. This integration, as it is further proposed, should enable a proper incorporation of the embodied dimension, including Gestalt principles, into theorizing and researching academic ELF. In this way, the many clines along which LUG and CxG are placed, all of which are essentially related to the emergent-emerging continuum, could be reconciled.

The second part of this dissertation is devoted to delineating a framework for the above-mentioned reconciliation, presenting Cognitive (Construction) Grammar (henceforth C(C)G) as an approach capable of compensating for the gaps (already) identified in usage-based theorizing of ELF. Thus, the third chapter synthesizes the cognitive and constructionist commitments of C(C)G. Firstly, the chapter presents Gestalt-based cognitive processes as uniquely highlighted in C(C)G and explains how these processes are connected to basic concepts, and thus foundational to the conceptual system and the representation of constructions. Next, the chapter elaborates on the notion of a construction and introduces three clines by which constructions can be described, i.e. fixedness-novelty, specificity-schematicity, and simplicity-complexity, simultaneously emphasizing that constructionhood can be observed at many levels.

The fourth chapter, in turn, aims to describe a C(C)G approach to constructions at one such level, i.e. discourse. The chapter begins by defining discourse and presenting discourse genres as constructions, highlighting their (potentially) multimodal nature, and then discusses ways in which discourse genres can be researched as constructions. To align with the division introduced in Chapter 2, the chapter presents selected studies on both spoken and written data, uncovering the heterogeneity of the solutions adopted and thus the many ways in which non-academic and academic discourse genre constructions can be interpreted. The chapter concludes that the uncertain ontological status of discourse genre constructions, which is particularly noticeable in the academic domain and in the written mode, requires further C(C)G-based research so that a clear and replicable method of identifying and describing discourse genre constructions could be devised.

The final part of this dissertation describes the study designed to develop a method of identifying and describing one type of discourse genre constructions, i.e. the elaborated abstract construction emerging in three academic ELF disciplinary communities. The fifth chapter presents the DISCOWER corpus, i.e. a new written ELF corpus created with the intention of providing a database for the identification of the elaborated abstract construction. Hence, the chapter presents how the assumptions of LUG and CxG are integrated in and through the cognitive and constructionist commitments characteristic of C(C)G to guide the process of compiling the DISCOWER corpus. To be more specific, in accordance with the general principles of corpus compilation, the creation of DISCOWER is described with reference to four aspects, i.e. corpus units, external criteria, corpus balance, and methods of data description. The chapter concludes by highlighting the importance of comprehensively characterizing elaborated abstracts tentatively delineated in the corpus as constructions.

Finally, the sixth chapter aims at systematically identifying and exploring the emerging elaborated abstract construction, which could be perceived as a stabilizing discourse genre construction attributed to academic ELF users. To identify and explore the construction, the continua of features through which constructions are described in C(C)G are used and hence elaborated abstracts are presented with reference to their degrees of fixedness, specificity and simplicity. Combining quantitative and qualitative methods, the elaborated abstract construction is reliably depicted as emerging in the academic ELF disciplinary communities of law, linguistics, and literary studies. The chapter concludes with a summary of its findings as well as presenting the limitations observed.

The dissertation ends with a conclusion that outlines potential implications of the study in relation to both the research fields embraced in the present work, i.e., academic ELF and C(C)G, as well as related areas of inquiry.

Chapter 1
English as a lingua franca: from language contact to language usage

1.0. Outline

The aim of Chapter 1 is to discuss the phenomenon of English as a lingua franca (henceforth also ELF). The chapter is divided into two main sections, characterizing ELF from a language contact perspective (Section 1.1.) and theorizing ELF: towards a usage-based approach (Section 1.2.).

Section 1.1. describes ELF from a sociolinguistic perspective, i.e. as emerging from language contact. The section commences with a brief description of how lingua francas can be classified and concentrates on three continua of features adapted from relevant classifications, i.e. a pidginized-natural continuum, a regional-global continuum, and a restricted-universal continuum.

Section 1.2. discusses two theoretical approaches to ELF, i.e. competence-based and usage-based, which arose alongside the language contact perspective with a view to highlighting the mechanics of language use and its relations to linguistic knowledge. Each approach is further divided into foci, or clines, i.e. deficient-different and emergent-emerging, respectively, to foreground the particular ways in which competence- and usage-based approaches are implemented in ELF theorizing.

The chapter ends with a summary which not only recapitulates the most important aspects of characterizing and theorizing ELF but also highlights areas that are worth further exploration.

1.1. Characterizing ELF from a language contact perspective

To delineate a theoretical framework for characterizing English used as a lingua franca, i.e. a common language among people without a shared mother tongue (Mauranen 2017), it seems necessary to begin with an elucidation of what the notion of language contact entails and what the terms Lingua Franca and lingua franca mean.

From a sociolinguistic perspective, the notion of language contact encapsulates interactions among people from heterogeneous linguistic milieus (Weinreich 1952; Milroy 2002; Seid 2015; Adamou and Matras 2021), which are seen as a driving force

behind language change (Dolberg 2019). In other words, when individuals who use different languages engage in communication, their interactions foster, among others, the exchange of words or grammatical structures between languages, which may gradually engender the development of a new language that facilitates communication among people without a shared mother tongue (Lindstedt 2009; Matras 2009; Baker and Matras 2013; West 2013).

One of such languages is Lingua Franca, i.e. a simplified language based on Romance languages which appeared as a tool for interethnic communication in the Mediterranean region (Mackenzie 2014; Formentelli 2017; Operstein 2022; Tweedie and Johnson 2022). According to Vikør (2004), the exact meaning of Lingua Franca remains unknown; however, the common belief is that "franca" can stand for either "Frankish" or "French"[1]. In other words, one of the hypotheses concerning its origin suggests that the term referred to the language used by the Portuguese or the Italians (commonly referred to as Franks), who played a pivotal role in maritime trade during medieval times. At the same time, another hypothesis implies that Lingua Franca can refer to the language used during a period of French imperial influence, which dates back to the 17th and 18th centuries.

In the following years, when Lingua Franca gradually lost its prominence and gave way to other languages, its name started to encompass a broader spectrum of languages that fulfilled a similar communicative function (Brosch 2015). Initially, the term was applied in relation to appearing new languages similar to Lingua Franca. Later, it also came to be employed with reference to existing native languages that were adapted to function as the bridge during international communication. As a result, the term Lingua Franca evolved into a label lingua franca and is now used in relation to any language, including English, facilitating communication without a shared mother tongue (Trudgill 2016; Gardner and Lau 2018; Bayyurt et al. 2019).

A detailed discussion of lingua francas is dependent on the classification adopted. For instance, Samarin (1978), Barotchi (1994, in Fiedler and Brosch 2022), or Back (2015) distinguish three types of lingua francas, i.e. planned, pidginized, or natural, reflecting their more or less artificial ways of development. A more fine-grained account

[1] An alternative explanation revolves around the idea that "franca" can denote "free", signifying a lack of specific ties of the language to any countries or ethnicities. However, as Vikør (2004: 329) comments, "this may be colored by a modern view on the desirability of freedom and equality between languages and their users"; hence, attributing this perspective to earlier centuries should be done cautiously.

is offered by Vikør (2004), who, apart from planned and natural lingua francas, proposes that lingua francas can be categorized as regional or global as well as specific to the domain of imperialism or the domains of religion and culture. In this way, lingua francas are presented in relation to their geographical range and the range of domains to which they apply. Likewise, Meierkord (2007) suggests that lingua francas can be classified as, among others, intranational or international and restricted or non-restricted to a specific domain. Simultaneously, the authors emphasize the potential of lingua francas to shift between categories as they evolve (Vikør 2004) and acknowledge that lingua francas are perhaps best depicted with reference to continua (Fiedler and Brosch 2022).

Thus, in the following sub-sections, the characterization of English as a lingua franca is presented via recourse to three continua constructed on the basis of features utilized in the above-mentioned classifications[2] (Samarin 1978; Barotchi 1994; Back 2015; Vikør 2004; Meierkord 2007), i.e. a pidginized-natural continuum, a regional-global continuum, and a restricted[3]-universal continuum.

1.1.1. ELF within a pidginized-natural continuum

As signaled above, lingua francas can be divided into two categories, i.e. pidginized and natural (Fiedler and Brosch 2022). Pidginized lingua francas, a.k.a. pidgins (Samarin 1978), are defined as non-native languages that emerge when communication is required but a shared language is absent (Sakoda and Siegel 2003; Vincente 2007; Yule 2016). Such languages can be discussed with reference to, for instance, pidgins that are currently extinct[4], e.g. Pidgin Delaware, which developed among African soldiers and British officers in the 17th century (Goddard 1997), or Ndyuka-Trio Pidgin, which emerged in the 20th century among speakers of Ndyuka, Tiriyó, and Wayana (Velupillai 2015). At the same time, pidgins can also be discussed with reference to languages that are still in use, e.g. Nauruan Pidgin English, which appeared as the result of interactions among speakers of Chinese and Melanesian Pidgin English in the 20th century (Siegel 1990; Tryon and Charpentier 2004).

[2] Since English is a natural language and by definition not artificially created, I concentrate on other lingua francas which appear spontaneously during natural interactions, i.e. pidginized (Fiedler and Brosch 2022).
[3] Meierkord (2007) applies the term "restricted" with reference to both a planned language, i.e. artificially created, and a language used in a specific domain, i.e. airtraffic control. However, in order to combine Meierkord's proposal (2007) with the one delineated by Vikor (2004), I introduce the distinction between restricted and universal lingua francas to describe the continuum between lingua francas utilized in a given domain or in different domains.
[4] According to Grenoble and Whaley (2005: 18), extinct languages are the ones with no remaining speakers.

Concurrently, pidginized lingua francas can also be described in relation to a set of processes affecting languages on which pidgins are based (Bakker 1995, 2009; Winford 2006; Crowley 2009; Al-Jasser 2012; Romaine 2017; Day 2019). For instance, in relation to the process of simplification, Sebba (1997: 37) claims that "the grammars of [p]idgins [are] characteristically less complex than the grammars of their source languages". At the same time, McIntyre (2020) states that

> [s]implification involves both regularisation and loss of redundancy. Regularisation means the process of making irregular forms regular - for example, pluralising mass nouns by the addition of an <s> inflection (…). Loss of redundancy involves the deletion of linguistic elements that repeat information. For instance, a pidgin might not use the third-person present tense inflection on verbs since this is grammatical information that is conveyed by the third-person pronoun.

Simultaneously, it should be noted that pidgins can undergo several stages of development, e.g. from jargons through expanded pidgins to creoles, thereby becoming gradually more complex (Walczyński 2012; Romaine 2017). The final stage of its transition entails that a pidgin begins to be applied as a native language (Watts 2011: 86). For example, Tok Pisin, which appeared as an English-based pidgin in the late 19th (Stroud et al. 1992), is now "the mother tongue of an ever increasing number of Papua New Guineans" (Tryon and Charpentier 2004: 10). On this account, when a pidgin becomes "enriched, expanded and regularized [i.e. a creole]; it has the full complexity characteristic of any natural language" (Gramley and Pátzold 2003: 460) and thus becomes similar to a natural lingua franca.

A natural lingua franca is defined as a mother tongue (Back 2015) which gains the role of a shared language among people without common linguistic backgrounds. For natural languages, such a role is solely an additional function, alongside the ones commonly associated with native languages, e.g. marking identity (Fiedler 2011). Natural lingua francas can be discussed, for instance, with reference to examples that are either dead[5] or still applied as native ones. An example of a dead lingua franca is Latin, which, up to the 16th century, was an extensively utilized common language, predominantly among the clergy (Back 2015). Simultaneously, among languages still applied as mother tongues, one can enumerate Spanish and Portuguese, which became dominant as lingua

[5] Trask (1999) claims that there are two types of dead languages, i.e. the one that disappears because there are no native speakers of this language and the one that disappears because "over a period of time it changes so substantially that its later forms are so different from the earlier form (…) that it no longer makes sense to apply the same name". Trask (1999) illustrates the second type with reference to Latin, which, although perceived as the base for the emergence of such languages as, for example, French, is seen as a dead language.

francas in the 16th century, Dutch, which gained prominence in the 17th century (Vikør 2004), or English, which has been an indispensable lingua franca since the mid-20th century.

When natural languages are employed as lingua francas, they can also undergo gradual simplification. As Back (2015: 10–11) states, the majority of "structural and grammatical changes can be traced back and grouped under the banner of simplification, a linguistic process that eliminates redundancy and ambiguity as far as possible". For instance, Björkman (2013) indicates that English used as a lingua franca exhibits tendencies similar to those found in pidgin languages. To be more specific, Björkman (2013: 149) claims that the reduction of redundancy can be noticed since "we see 'not marking the plural on the noun', 'subject-verb agreement issues' and the 'non-standard analytic comparative'". However, if such alterations of natural languages become significant, a natural lingua franca is deemed to have changed the category; thereby functioning as a pidgin (Samarin 1978; Back 2015).

1.1.2. ELF within a regional-global continuum

Over the centuries, both pidginized and natural lingua francas have occupied different geographical areas, functioning as (more) regional or global shared languages. According to Vikør (2004: 333), regional lingua francas "cover either a certain geographic area or a group of related countries in different areas". This definition, in turn, implies a lack of a precise demarcation of what a regional lingua franca is since its use can be, for instance, national, continental or intercontinental. To illustrate this point, Nigerian Pidgin, as the name suggests, is attributed to Nigeria (Todd 2005), where it functions as a national lingua franca. Continental usage can be described in relation to German, which is utilized as a lingua franca in such European countries as Switzerland, Austria, and Germany (Meierkord 2006). Intercontinental usage can be discussed with reference to Portuguese, which is used as a common language in both Europe and South America (Meierkord 2006). At the same time, the dynamic nature of lingua francas further complicates the establishment of a precise regional delineation as lingua francas can expand their geographical scope over time. For example, Sanskrit, initially dominant in India, gradually spread beyond the boundaries of this country (Vikør 2004).

At the uppermost level of the aforementioned scale, the "global" dimension signifies the widespread use of a lingua franca around the entire world. This role, in fact, is attributed solely to one language, i.e. English, as it "has spread around the globe like

no other language before" (Mauranen 2018). Its global prominence, however, was neither a rapid occurrence nor a simple coincidence. Instead, the emergence of English as the world's lingua franca is the culmination of a complex interplay of various factors that, over centuries, have contributed to its prevailing status. In other words, "although the global spread of English as a lingua franca belongs very much to the present, it needs to be put in perspective by reference to the past" (Seidlhofer 2011: 11).

The geographical expansion of English is discussed in detail by Gómez (2020), who claims that it traces back to the 16th century when English first arrived in North America. Subsequently, in the 17th century, it appeared in the region of the Caribbean and, at that moment, its gradual expansion was mostly associated with colonization. As the 18th century approached, English spread throughout Southeast Asia, becoming the language of law and administration in this region. At the same time, both Australia and New Zealand experienced the expansion of English, stemming from the establishment of a convict colony in Australia and a steady influx of immigrants in New Zealand. However, in the 18th century, English continued to spread due to not only colonization but also the Industrial Revolution, which originated in Great Britain. Hence, English started to be seen as a key to new technological developments and mastery of the English language became synonymous with progress and prestige.

In the 19th century, the diffusion of English continued, reaching certain parts of the African continent and leaving its mark in the South Pacific. However, it was after World War II that it accelerated significantly. This rapid expansion was closely intertwined with globalization, driven by business-related issues and technological advancements. Nowadays, as Pontello claims (2020: 4), "[t]he recent phenomena of globalization and digitalization have enabled the English language to spread globally and to proliferate across every social sector of the world and across many countries of the planet". Admittedly, since globalization is perceived as continuously increasing (Capozza 2023), the number of ELF users, which has already surpassed the number of native ELF speakers (Sung 2014; Wu et al. 2019), is still expected to rise (Brunner 2022).

1.1.3. ELF within a restricted-universal continuum

Apart from a geographical perspective, the development of lingua francas can also be examined in terms of domains in which they are utilized, giving rise to more restricted and universal lingua francas. As noted by Meierkord (2007), "[s]ome lingua francas serve highly specialized purposes and function only in very restricted contexts such as airtraffic

control. However, others are employed for government purposes, as medium of instruction in schools and tertiary education (Smit 2003) and (…) intimate, personal interactions". Hence, a restricted lingua franca can be perceived as a lingua franca typically attributed to a particular domain. In contrast, a (more) universal lingua franca can be broadly defined as a lingua franca connected to more than one domain as there seems to be no precise indication of the number of domains leading to its universality.

For instance, Fanakalo, i.e. a pidgin which emerged in the South African mines, "is typically used in the work domain" (Mesthrie 2009: 265). Therefore, it can be seen as a restricted lingua franca that has not expanded the number of domains in which it is applied. Tok Pisin, however, which arose in the domain of trade (Schulte-Schmale and Naujoks 2013: 4), is described as "expanding more and more and is used in a wide range of domains", thereby functioning as a more universal lingua franca. Likewise, Latin, which is said to have emerged as a lingua franca due to political needs (Vikør 2004), is described as a shared language for religion, academia, culture, law, and administration (Salverda 2018), also showcasing an expansive nature.

As for English, taking into account the multitude of factors that have contributed to its unique global dominance, it can be seen as a lingua franca that has greatly expanded the range of domains in which it is used. In other words, from a lingua franca applied due to political exigencies in the 16th and 17th centuries, English has transformed into a universal lingua franca that is associated with a great number of domains. Björkman (2013), for instance, states that English is utilized in such domains as, among others, international organizations and conferences; scientific publications; international banking, economic affairs and trade; advertising for global brands; audio-visual cultural products; international tourism, or tertiary education. At the same time, it is acknowledged that the use of ELF is in fact dominant in the domain of business and academia (Mauranen 2010; Cogo 2015). Nevertheless, given its ongoing development, such a list is perhaps still expandable.

All in all, with reference to the three continua discussed above, English can be described as a natural language which functions as the global lingua franca in various domains, e.g. business or academia. Throughout history, numerous factors have contributed to the lingua franca use of English, including colonization, technological advancements, and, more recently, globalization and digitalization. Consequently, ELF has become a remarkably complex phenomenon and it is expected that its complexity will persistently transform with its ongoing progress (Jenkins 2015), constituting a challenge

"in terms of theorisation, description, and representation" (Baird et al. 2014: 1), or "the beast to be defined (…) described, and researched" (Pitzl 2018: 10). In the following sections, I thus concentrate on an overview of ELF research, illustrating the ways in which the global lingua franca has been theorized.

1.2. Theorizing ELF: towards a usage-based approach

As mentioned above, English has served as a natural lingua franca since the 16th century, gradually becoming the global common language. As a subject of scholarly inquiry, however, ELF garnered attention in the 1980s, with a substantial surge in research from the early 2000s (Gómez 2020). Throughout the years, scholars have approached ELF from various perspectives (Björkman 2013), "drawing from many different areas of linguistics and adopting a wide variety of theoretical and methodological perspectives" (Yilmaz 2020: 5). For instance, characterized as a contact language, i.e. a primarily social phenomenon, ELF, has, expectedly, been discussed within the World Englishes model (Pitzl 2016)[6]. At the same time, it has equally successfully been viewed from the perspective of, e.g., "politeness theory, (…) conversation analysis, ethnography, computer-mediated communication, and corpus analysis" (Millot 2015: 1). Moreover, given the constantly growing complexity of ELF, it is only natural "that theorising, too, cannot stand still, but that ELF scholars need to be ready to revise their conceptualisations of the phenomenon in line with new empirical findings as well as by considering conceptualisations and empirical research from other fields of enquiry that can 'speak' to ELF" (Jenkins 2015: 51). Two such "voices", i.e. a classical approach and a usage-based approach, have been identified by Vetchinnikova (2015).

To be more specific, it has been proposed that theories regarding ELF can be divided into two approaches, i.e. classical and usage-based. A classical approach assumes the existence of the abstract linguistic system that functions independently of language use, which, for Vetchinnikova (2015: 225–230), complies with the "Saussurean tradition with its notions of langue and parole as well as Chomskyan distinction between competence and performance"[7]. Conversely, a usage-based approach is described as the

[6] For a discussion of ELF in relation to, among others, pidgins and creoles see Björkman (2013) or Watkowska (2020).

[7] The notion of langue refers to an idealized system of rules which underlies (but is not fully mirrored in) specific instances of language use (i.e. parole) (Saussure 1959 [1916]). Hence, it is similar to the notion of competence, which is defined as idealized knowledge of a given language which gives rise to (but is not reflected in) language use (i.e. performance) (Chomsky 1965). However, since ELF researchers who refer

one which stresses that "language use is the only form in which language exists and so grammar [i.e. linguistic competence] is a description of language in use". In essence, as Vetchinnikova (2015: 8) puts it, the difference between these two approaches is motivated by a fundamental question that determines the way ELF is theorized, i.e., "[i]s there an abstract linguistic system/competence and does it underlie all usage/performance? Or is competence merely an abstraction from all instances of usage/performance?". The details of the above-mentioned perspectives are provided below.

1.2.1. A competence-based approach

As signaled above, a competence-based approach to ELF aligns with a "rather generativist (Chomsky, 1965) approach" (Yilmaz 2020: 13)[8], which stresses that cognition should be perceived in a modular way. As explained by Taylor (1995: 16), such a perspective entails that "[j]ust as the human body consists of various parts, each with its own function and developmental history, so the human mind consists of components which, though interacting, nevertheless develop and operate independently". One of the components is the language faculty, i.e. a mental module allowing for language production and comprehension (Chomsky 1980), which can be perceived in a broad or a narrow way.

The Broad Language Faculty is defined as all the abilities necessary for the acquisition and application of language, which are partially shared by other species, e.g. "human auditory, motor, and vocal systems, short- and long-term memory, and (joint) attention, among others" (Jackendoff 2011, in Dąbrowska and Divjak 2015). Therefore, such a view takes into account the interaction of the language faculty with other mental abilities. In contrast, the Narrow Language Faculty refers to a unique human predisposition to acquire and use language, entailing the existence of innate linguistic knowledge. As elucidated by MacKenzie (2014: 40), the Chomskyan approach assumes that the language faculty contains information which is "biologically determined, 'hardwired' in the brain [i.e. Universal Grammar (UG)] without which language would be unlearnable". Much in the same vein, Cowie (2008) claims that

> [o]n Chomsky's view, the language faculty contains innate knowledge of various linguistic rules, constraints and principles; this innate knowledge constitutes the 'initial state' of the language faculty. In interaction with one's experiences of language during childhood — that

to the classical approach point to its convergence with the Chomskyan model (Seidlhofer 2011; Alptekin 2013; Yilmaz 2020), I adopt an analogical perspective and thus re-label the approach as competence-based.

[8] The alignment does not indicate that the theories developed in ELF adhere to all the assumptions of the generativist approach. In fact, the main focus is put on the ideal nature of linguistic knowledge as well as the inherent capacity of the language faculty to generate an infinite number of sentences.

> is, with one's exposure to what Chomsky calls the 'primary linguistic data' or 'pld' (…) — it gives rise to a new body of linguistic knowledge, namely, knowledge of a specific language (like Chinese or English).

In other words, the process of learning involves scanning "the ambient linguistic input (…) to which the child is exposed and set[ting] the predefined parameters of UG to match it" (MacKenzie 2015), leading to competence in a given language[9].

However, "the linguistic input children receive is of too poor a quality to explain the quantity and quality of their linguistic output after the age of two" (Chomsky 1965, in Mackenzie 2014). Therefore, the generativist perspective (Chomsky 1965) assumes that the process of language learning transcends mere imitation of linguistic input. In essence, building on the idea that, for instance, children are able to produce and comprehend linguistic expressions never uttered by their parents, Universal Grammar is believed to inherently facilitate the innovative exploitation of linguistic resources, enabling humans to construct sentences that they have not encountered before. In other words, as Pinker (2003) states, "the brain must contain a recipe or program that can build an unlimited set of sentences out of a finite list of words". Hence, Universal Grammar is viewed as a "generative device" or a "computational system" (Gärtner 2014: 14).

Simultaneously, a distinction between linguistic competence and performance, i.e. language use, is highlighted. As Chomsky (1980: 224) claims, language use is influenced by other mental factors such as, for example, memory limitations or shifts of attention (Chomsky 1965: 3). Chomsky (1965: 3–4) clarifies this assumption by noting that a transcript of natural speech often reveals a multitude of false starts, deviations from established rules or abrupt changes in expressions. Therefore, even though linguistic competence underlies performance, performance cannot fully reflect linguistic competence (Evans and Green 2006). For a language learner or a linguist, the challenge is thus to discern the underlying system of rules mastered by speakers "in virtue of the properties of the language faculty" (Taylor 1995).

Consequently, the Chomskyan approach is perceived as "equat[ing] language with (or reduc[ing] it to) grammar alone" (Mackenzie 2014: 40). This assumption, at the same time, entails that the approach is attributed to a formalist perspective (Mackenzie 2014), i.e. the one that "looks on language as a formal object, defined by rules" (Seuren 2004: 98). In other words, it is assumed that little attention is given to the role of communication

[9] To illustrate, when people learn English, they do not learn the concept of sentence subject. Instead, learners discover that the notion of sentence subject is in English conveyed through word order and thus a link between English and innate linguistic knowledge is established (Pinker 1984).

and the meaning of forms through which linguistic competence is described. Moreover, such a perspective posits that the Chomskyan approach centers on the autonomy of the language faculty and thus extreme idealization (Taylor 1995). On this account, Grenfell and Harris (2002: 16) state that "Chomskyian linguistics is concerned with the ideal competence" of an individual, i.e. "an ideal speaker-listener (…) who knows its language perfectly" (Chomsky 1965: 3).

In ELF, a competence-based approach is linked with two perspectives: a focus on deficient ELF (Section 1.2.1.1.) and a focus on different ELF (Section 1.2.1.2.), each with its own interpretations of ELF and its users.

1.2.1.1. A focus on deficient ELF

A focus on deficient ELF is closely associated with the concept of native-speaker competence, which involves the capacity for generating linguistically "correct" forms (Seidlhofer 2017: 93). As Seidlhofer (2017: 93) notes, it thus aligns with Chomsky's perspective on linguistic knowledge, which entails an ideal individual, who possesses perfect knowledge of his or her language (Chomsky 1965: 3). In this focus, the idea of correctness or perfection is frequently associated with the notion of Standard British and American English (Melchers et al. 2019; Dewey and Pineda 2020), which serve as the models to be followed (Seidlhofer 2008; Grazzi and Lopriore 2020).

Standard British and American English hold the status of varieties which have garnered widespread recognition in grammar books and dictionaries. With sets of codified norms, these models are commonly regarded as accurate and stable (Trudgill and Hannah 2017). As a result, their significance goes beyond being just ideal examples. In fact, their enduring dominance serves as a crucial reference point in teaching English as a foreign language, playing a role in creating a sense of global linguistic homogeneity. Simultaneously, their significance also plays a role in shaping the perception of ELF. In other words, any deviation from the norms of native speakers, which is typical in the ELF context (Seidlhofer 2017), is perceived as inaccuracy.

ELF is thus defined as, for instance, "a deviation from the Standard English" (İnceçay and Akyel 2014, in Asakereh et al. 2019) or interlanguage (Gomez 2020), i.e. "learners' versions of the target language" (Hu 2016: 63). Concurrently, in line with the notion of interlanguage, individuals engaging in international communication are perceived as learners of English as a foreign language, who are in the process of approaching language proficiency comparable to that of ideal native speakers (Gomez

2020). On this account, scholars following this focus question the legitimate status of ELF and highlight that, even though ELF holds a dominant worldwide presence (see Section 1.1.2.), learners make efforts to follow the conventions set by native speakers.

For example, one of the critical views suggests that attention paid to ELF could result in the development of a framework based on the principle of "anything goes" (Jenkins 2006: 141). As Davies (1991, 2003) notes, the absence of the traditional native-speaker model as guidance for English learners places them in an unfamiliar realm without clear points of reference. Likewise, Sobkowiak (2005: 141), in the context of ELF perspective on pronunciation, claims that ELF could "bring the ideal down into the gutter with no checkpoint along the way". Simultaneously, another critical argument is that there is no need to create new models. For example, Trudgill (2005) posits that investigations into ELF can be seen as an endeavor to replace one standard with another, which seems questionable if such models as Received Pronunciation or General American are already available. A similar perspective is presented by O'Regan (2014), who claims that ELF scholars inadvertently attempt to impose standards for learners to follow. This approach, O'Regan suggests, perpetuates a form of paradox, where scholars ostensibly advocate for the empowerment of underprivileged individuals, i.e. non-native speakers of English, while simultaneously curbing their agency in deciding which variant of English to employ. Furthermore, drawing upon empirical research, Kuo (2006) states that a strong inclination to adhere to native-speaker norms remains prevalent among learners, extending beyond those who use or intend to use English primarily with native speakers (Timmis 2002: 248). In a similar vein, Groom's study (2012) corroborates Kuo's (2006) assertion, as Groom underscores that English learners tend to resist embracing ELF due to their negative attitudes towards being recognized as non-native speakers.

Nonetheless, as Widdowson (2016) suggests, even though Standard English and native speakers are useful concepts that offer a certain degree of stability, they are merely privileged notions. A similar view is expressed by Saraceni (2015: 7), who claims that "to the vast majority of laypeople the idea that 'it is no longer accurate to say that there is just one English' is still controversial". Hence, while British and American English are portrayed as well-known varieties, it is emphasized that there are actually numerous alternatives, thereby leading to a focus on different ELF.

1.2.1.2. A focus on different ELF

The adoption of a focus on different ELF stems from an effort to describe ELF as "different rather than deficient" (Jenkins 2015), building on the idea that "different types of language use, whether ELF or ENL [i.e. native English], are alternative actualizations of the same abstract system of rules" (Yilmaz 2020: 13). In other words, a focus on different ELF aims to describe the competence of ELF speakers with reference to the virtual language, i.e. "abstract linguistic resources, reminiscent of Chomsky's generative rules that are biologically preset" (Alptekin 2013: 201) or, to put it differently, a generative potential for variable realizations (Widdowson 1997, 2016).

According to Seidlhofer (2011), the concept of the virtual language can be elucidated with reference to two types of rules, i.e. constitutive and regulative. Constitutive rules (the virtual language) serve as the foundational encoding possibilities that constitute the system and determine its boundaries. As Widdowson (2016: 33) emphasizes, "contrary to what has at times been supposed (…) the concept of a virtual language is not a system of actual encodings". In fact, such rules are seen as similar to "the system of generative principles and constraints (…) in the Chomskyan sense" (Paradowski 2013: 313). Regulative rules, in turn, are described as context-dependent conventions concerning the application of constitutive rules (Seidlhofer and Widdowson 2009). For Widdowson and Seidlhofer, such rules are noticeable in data on linguistic performance, i.e. corpora, which are simultaneously perceived as "reductionist" (Seidlhofer 2011: 113), i.e. providing "only a partial account of the real language" (Widdowson 2000: 7).

The distinction between constitutive and regulative rules is incorporated into ELF theories in order to emphasize its legitimate status. As Seidlhofer (2011: 114) notes, the perception of native-speaker norms as the only "real language" entails treating a specific set of regulative conventions as constitutive rules. In other words, such an approach presumes not only that the competence of a native speaker is the best one but also that it is the only one. Thus, when attention is shifted from the regulative rules to the constitutive rules of virtual English, one can state that native English, English as a lingua franca, or other Englishes are, in fact, legitimate realizations of the same underlying system.

For instance, according to Widdowson (2016), Standard English is just one of the possible realizations of the virtual language that has been conventionalized within a specific community. As Widdowson claims (2020: 33),

> [t]his is only one version of the language, a variety that has been accorded a privileged status but of its nature not essentially different from any other. Like the others it is the exemplification of certain encoding principles, one set of realizations that have become conventionally established within a particular community. But a code is of its nature a generative device with the potential to be realized in all manner of various ways.

To illustrate, Widdowson (2020) states that the virtual language permits the application of the plural suffix to all nouns; irrespective of their mass and non-mass status. Therefore, even though words like "informations", "evidences", and "advices" are not commonly used in Standard English, and are regarded as "errors" (Seidlhofer 2011: 15), the words comply with the underlying principles of the virtual language. In a similar way, the utilization of nouns like "window" and "door" as verbs is equally consistent with the rules of the virtual language as the employment of nouns like "table" and "chair" as verbs; the only distinction is that the latter have already been coded.

Hence, when ELF users (or EFL learners) employ the resources in an unconventional way, they may be said to incorrectly follow the encoding rules. Nevertheless, "the rules do apply: it is just that users of English have hitherto not had occasion to apply them, or are inconsistent in their application" (Widdowson 2020: 34). Based on this premise, a focus on different ELF attempts to find regularities concerning the use of ELF in the global setting, which leads to the assumption that ELF is also potentially codifiable (Jenkins et al. 2011).

One of the most popular lists of ELF features is the outcome of the study by Seidlhofer (2004), who proposes the general lexicogrammatical tendencies of ELF. To be more specific, Seidlhofer (2004: 220) states that the use of ELF can be characterized by:

> 'Dropping' the third person present tense –s; 'Confusing' the relative pronouns who and which; 'Omitting' definite and indefinite articles where they are obligatory in ENL, and inserting them where they do not occur in ENL; 'Failing' to use correct forms in tag questions (e.g., isn't it? or no? instead of shouldn't they?); 'Inserting' redundant prepositions, as in We have to study about...; 'Overusing' certain verbs of high semantic generality, such as do, have, make, put, take; 'Replacing' infinitive-constructions with that-clauses, as in I want that; 'Overdoing' explicitness (e.g. black color rather than just black).

As Seidlhofer emphasizes (2004, 2011), although such forms do not comply with the norms of Standard English, people using ELF do not have difficulty in achieving mutual understanding. Seildhofer (2011) explains this fact by stating that successful ELF interactions are based on the exploitation of the virtual language, which is strategically adapted to the needs of a particular communicative situation. As a consequence, Seildhofer (2011) indicates that the notion of an English learner, which implies a deficiency in the knowledge of Standard English, should be replaced with the notion of

an ELF user, which emphasizes a proficiency in coping with the intricacies of ELF communication.

However, Alptekin (2013) notes that connecting the virtual language with the ability to utilize resources tailored to a specific context centers solely on highlighting the ELF context. As Alptekin (2013: 201) claims, "little is offered as regards the complex and dynamic interrelationship between users' interactions with the social context and their exploitation of their cognitive resources, with a view to generating contextually driven linguistic forms that function to communicative effect". According to Alptekin (2013: 201), efforts to explore ELF with a "convenient fiction at an almost Chomskyan level of abstraction" do not provide a fine-grained account of the meaningful relationship between form and function as merely the strategic competence of ELF speakers is emphasized. Much in the same vein, Hall (2017) states that the notion of the virtual language seems rather imprecise and acknowledges the need for alternative approaches to its definition (see Section 1.2.2.2.), which leads to the introduction of a usage-based approach to ELF (Alptekin 2013; Hall 2017).

1.2.2. A usage-based approach

As noted by Butler and Gonzálvez-García (2014: 2–6), "usage-based linguistics" is one of the terms that are "potentially misleading and are used in different ways by different scholars". Originally, the term "usage-based" was coined by Langacker (Behrens 2009) to "highlight a methodological and theoretical contrast between cognitive and generative linguistics" (Mengden and Coussé 2014: 2). More specifically, the notion was proposed in a narrow sense, i.e. to distinguish Langacker's model of Cognitive Grammar (1987, 2005) from Generative Grammar (Chomsky 1988, 2013) and support the former's assertion that linguistic knowledge emerges from language use and "should be studied as a reflection of cognitive processing and hence a source of information about cognitive abilities and mental phenomena in general" (Boers 1996: 13). Over time, the scope of the term has expanded and, currently, it "can also be used in a broader sense to refer to a family of approaches which share similar views on the relationship between linguistic knowledge and language use" (Dąbrowska 2008: 4) on the one hand, and language and cognitive processes, on the other hand. This family of approaches encompasses, e.g. Emergentism, Complex Dynamic Systems Theory, or Construction Grammar(s), all of which can be characterized with reference to a common set of assumptions (see, for instance, Barlow and Kemmer 2004; Evans and Green 2005; Dąbrowska 2008; Behrens

2009; Bybee 2010, 2023; Ibbotson 2013; Baird et al. 2014; Diessel 2015, 2017; Hall 2017)[10].

To begin with, a usage-based approach rejects the existence of innate linguistic knowledge. As noted by Baird et al. (2014: 184), linguistic knowledge is "not a pre-given entity with a set of forms and rules, but it is something constructed through people's linguistic communication [i.e. performance]". According to Schmid (2020: 15), communication occurs in the form of specific instances of language use through, among others, speaking, handwriting, or typing. Therefore, Evans and Green (2005: 109) claim the most important role within a usage-based approach is to be attributed to "a situated instance of language use which is culturally and contextually embedded".

Such instances of language use involve, among others, a specific choice of words or phrases (linguistic communication) or the interplay of words and intonation or gestures (linguistic and para/non-linguistic communication)[11] to convey a meaning shared within a community[12]. The motivation behind particular linguistic (and paralinguistic) choices is adaptation to the context of a given interaction, which encompasses, among others, participants (their origin or age) or a situational type in which the participants are engaged (a casual chat or a job interview) (Schmid 2020). To clarify, a teacher may communicate with his or her students by choosing specific words or phrases pertinent to an academic lecture to emphasize key information. Such form-meaning/function combinations constitute the ground from which linguistic knowledge emerges under the influence of cognitive processes.

[10] Another option is to adopt an even broader perspective, i.e. such that embraces both functional and cognitive-functional approaches. This perspective is adopted by, for instance, Tyler (2010), who applies the term usage-based in relation to, for instance, Systemic Functional Linguistics (functional) and Cognitive Linguistics (cognitive-functional). Although this broader understanding offers an invaluable opportunity to rely on "different but complementary empirical and theoretical approaches" (Ellis et al. 2016: 24) under the umbrella term "usage-based", it seems to overlook the essence of a usage-based approach, i.e. the fact that "language emerges not as a result of a language-specific instinct but through the interaction of cognition and use" (Ibbotson 2013: 1). As noted by Alptekin (2013: 201–202), a usage-based approach is set apart from a functional approach "in terms of its sociocognitive commitment to developing a model of language use and knowledge that takes into account both the language user's apprehension of the social context and activation of the underlying cognitive network". Sharing Alptekin's (2013) view, I define a usage-based approach accordingly.

[11] While from a fine-grained perspective language use may indeed involve linguistic, paralinguistic and non-linguistic components (Ungerer 2023), from a coarser-grained perspective, advocated by Cienki (2015) and adopted in this dissertation, a distinction into linguistic and paralinguistic elements is sufficient.

[12] As noted by Hall (2017: 76), "[u]sage-based linguists generally assume a supra-mental ontological category of language that holds at the community level, and in fact most work within the approach is concerned with group knowledge of national varieties". This implies that the term community is prototypically linked in a usage-based approach with the notion of a speech community which is defined as a group of people with shared linguistic units within a certain geographical space (Saraceni 2015).

Since a usage-based approach prioritizes a non-modular view of cognition, the emergence of linguistic knowledge from language use is deemed possible due to domain-general cognitive processes. For Bybee (2010), for example, there are five cognitive abilities that facilitate the development of linguistic knowledge, i.e. 1) categorization – the recognition of instances as examples of a specific category, 2) chunking – the creation of units through frequent use, 3) rich memory – the retention of such units, 4) analogy – the application of existing units to novel instances, and 5) cross-modal association – the ability to connect form and meaning. Concurrently, Hall (2016) delineates three groups of cognitive constraints. The first group entails domain-general cognitive constraints, e.g. the brain's natural structure that facilitates information processing. The second group stands for cognitive skills allowing for interaction with other people, e.g. the ability to understand intentions of others. The third group refers to pattern-findings skills, e.g. the ability to establish schemas based on recognized similarities between instances. Likewise, Diessel (2017) claims that there are three groups of cognitive processes, i.e. 1) memory and processing, involving, among others, automatization (leading to easier activation of units) or priming (facilitating the activation of linguistic knowledge due to prior related information); 2) social cognition, referring to a set of processes enabling interactions, e.g. joint attention (focusing attention on the same experience) or a common ground (establishing a shared knowledge base to interpret information); and 3) conceptualization, entailing construal operations (see Section 3.1.1.2), e.g. Gestalt-based processes of grouping and figure-ground segregation (Diessel 2017: 10).

Admittedly, although the above classifications tend to include processes that differ in number and/or are referred to by means of different terms, they generally point to the continuum between individual and collective processes. In other words, some of the above-mentioned processes involve abilities that need not entail interpersonal interactions, e.g. categorization, while some are closely related to them, e.g. establishing a common ground. On this account, the appearance of linguistic units, a.k.a., for instance, constructions (Goldberg 2006), cognitive routines (Diessel 2017), or resources (Canagarajah 2020), can be described with reference to (the gradation between) individual and collective levels, standing for the linguistic knowledge typical of a particular person or shared within a given community, respectively.

` At the same time, linguistic units can be discussed from a more temporary or a more stable perspective. This idea aligns with Breyer et al.'s (2011) division into two

complementary approaches, i.e. emergent and emerging[13]. The former involves "considering language in the process of its usage at a given moment according to the situational and processual dynamics of a particular conversation" (Breyer et al. 2011: 193)[14], while the latter focuses on the relative stability of units resulting from their frequency of occurrence beyond the confines of a particular situation. To be more specific, as Breyer et al. (2011: 193–194) elucidate,

> a) Emergent theorists explore how (fragments of) form-function units pass from talk to talk, presuming that grammar exists as (many) grammars, differing according to communicative genres or activity types. Emerging theorists, on the other hand, explore the cognitive foundation of constructions, presuming that frequency leads to entrenchment via processes such as analogy.
> b) Emergent theorists concentrate on the process whereby new combinations of (previously heard) forms are made in interactive encounters. They want to know how speakers deal with (old) fragments. Emerging theorists, in contrast, are interested in the result, i.e. new form-function units, newly categorised elements and re-organised systems. They aim to describe and explain (new) categories.
> c) Emergent theorists are intrigued by how speakers create grammar as they go along and how they use linguistic resources in concrete situations. They thus focus on the present moment (…). Emerging theorists, on the other hand, explore languages as self-organising, complex systems and sometimes postulate an a priori grammar.

In the realm of ELF, the adoption of a usage-based approach is linked with the appearance of both perspectives, i.e. emergent and emerging. In other words, a focus is either on describing ELF as temporary and situation-bound (Section 1.2.2.1.) or on capturing its stabilizing character beyond a particular interaction (Section 1.2.2.2.).

1.2.2.1. A focus on emergent ELF

Apparently, an emergent perspective is said to best reflect the nature of ELF, which is often depicted as "one-off, transient, fleeting, and highly emergent" (Jenkins 2017: 11) or "elusive, ever-evolving, and dialogic" (Prodromou 2007: 411). As Jenkins explains (2017: 11),

> [w]ith the increasing spread of English among an ever-wider range of language groups, and massive growth in global mobility, the potential for diversity in ELF communications is itself

[13] Although it is convenient to distinguish between emergent and emerging perspectives, especially for expository purposes, there are reasons for "a reconciliation between emergent and emerging" perspectives (Auer and Pfänder 2011: 15), i.e. for a mixed, or an integrated approach, through which "[t]he opposition of emergent and emerging constructions can be overcome" (Auer and Pfänder 2011: 18). Simultaneously, this cline can be complemented with the emerged perspective, i.e. the one that focuses on already established units. In other words, the development of linguistic knowledge can be presented in relation to ad hoc units (i.e. emergent), stabilizing units (i.e. emerging), and stable units (i.e. emerged) functioning as "old". This perspective is, in fact, developed in the following chapters.

[14] Although a conversation is a canonical example of an interaction, online encounters between authors and readers of texts are also viewed as interactions (Mauranen and Sinclair 2006).

> becoming ever-greater, and transient encounters among ELF users is arguably nowadays more typical of ELF communication than interactions entailing the longer- or even shorter-term 'groups' that have been the primary focus of ELF research to date.

Consequently, because of ever-changing linguistic and cultural backgrounds that are (potentially) present in ELF interactions, Canagarajah (2007) stresses that it "is difficult to describe this language a priori. It cannot be characterized outside the specific interaction and speakers in a communicative context". Likewise, Meierkord (2004: 129) claims that ELF "emerges out of and through interaction"; therefore, "it might well be that ELF never achieves a stable or even standardized form". Hence, Kimura and Canagarajah (2017: 301) claim ELF interactions assume "diversity as the norm, and hence the starting point of interactions".

In other words, "[w]hat is shared (…) may not be shared from the start (nor do interlocutors necessarily know from the start what they do in fact 'share'). This means that we are often talking not of a priori 'resources' but of resources that are discovered as they emerge during the interaction" (Jenkins 2015: 64, in Jenkins 2017: 11) during which ELF users "negotiate with the parties and resources in a setting for constructing meaning" (Canagarajah 2020: 305–306). In fact, these resources may be "so deeply intertwined and fused into each other that (…) it [might be] difficult to determine any boundaries that may indicate that there are [for instance] different languages involved" (Makoni and Pennycook 2012: 447).

In line with the above, Jenkins (2015) states that ELF communication should be perceived as involving "repertoires in flux", i.e. resources that are constantly shaped and utilized in a way that aligns with the needs of a given communicative situation[15]. At the same time, given the emphasis on the temporary character of ELF interactions, the traditional notion of a community, which entails the existence of stable linguistic knowledge among its members (Kecskés 2014), is replaced within a focus on emergent ELF. One of the solutions is proposed by Pitzl (2018), who introduces the concept of transient international groups, i.e. temporary groups consisting of multilingual ELF users who come together for a specific purpose in a particular location for a limited duration. Thus, ELF users are seen as interacting here and now, without developing stabilizing units. Likewise, Jenkins (2015) proposes to rely on the notion of "contact zones", which

[15] As noted by Godwin-Jones (2020: 9), instead of the notion of competence, the idea of "communicative repertoires has been widely embraced in ELF".

is said to align with the character of "the temporary meeting and mixing of people from diverse backgrounds" (Mauranen 2018: 11).

Simultaneously, a focus on emergent ELF is criticized for paying too much attention to the temporary character of ELF. As noted by Mauranen (2018), for instance, there are certain ELF groups which can be seen as longer-lasting and thus able to develop stabilizing ELF units which can be felicitously captured through an emerging perspective.

1.2.2.2. A focus on emerging ELF

As signaled above, a focus on emerging ELF aims to provide an account of ELF beyond the confines of a particular communicative encounter, emphasizing individual and collective preferences of ELF users leading to the emergence of stabilizing ELF units.

To begin with, Alptekin (2011: 158–159) argues that "ELF users can be said to be developing their mutually intelligible and comprehensible forms based on the frequency of the output they produce and the input they receive from other ELF users as part of what can only be accounted for as both an emergent and emerging novel linguistic code". This "emerging and partially emergent linguistic code" comes with its own norms, which align with the requirements of the sociocultural contexts but which do not have to align with the norms of the native English.

Likewise, Mauranen (2012, 2017) claims that the linguistic knowledge of ELF users is less automatized and hence less convergent with native norms than the linguistic knowledge of native users of ELF. As a consequence, ELF users tend to apply "rough equivalents" of the target (native) units. For instance, units in ELF tend to have a preposition or an article different from their equivalents in native English. Therefore, Ranta (2013: 33–34) states that

> [o]n account of the fact that L2 speakers have had less exposure to English during their lifetime than L1 speakers, it is reasonable to assume that language forms are not as deeply entrenched in L2 users' minds as in L1 speakers'. (...) As the less deeply entrenched language forms require more effort in retrieving and processing in L2 speech, L2 users may start resorting to processing shortcuts and thus only approximating the target forms.

Approximation functions as an umbrella term which entails such processes as simplification, regularization, analogy, or generalization (Mauranen 2017, 2021, 2023). For example, the process of simplification can be illustrated by the tendency of ELF speakers to use mass nouns as countable nouns since countability follows a more predictable pattern than uncountability (see Section 1.1.). As a result, maintaining a

uniform plural marking is cognitively more transparent and it facilitates comprehension for both speaker/writer and hearer/listener.

At the same time, Mauranen (2019) notes that approximation, such as the omission of morphemes or articles, goes beyond a single communicative situation. Therefore, what becomes evident is a more intriguing inclination among ELF users to develop similar preferences across various interactions and linguistic backgrounds. Building on the proposal by Vechninnikova (2014), who claims that it is possible to delineate emerging preferences in ELF use among individual users (i.e. fixing), Mauranen further (2023: 31) suggests that recurring preferences among ELF users may "have repercussions for (...) communal language", i.e. they may lead to "collective fixing" (Mauranen 2017: 19).

Hence, with reference to the collective level, a focus on emerging ELF highlights that ELF can be described through the regularities pertaining to what is actually shared by ELF users. For instance, Mauranen (2017: 14–15) claims that ELF interactions involve reliance on highly frequent linguistic units and explains that

> when [two] speakers look for the least common denominator that would support interactional fluency, it is likely that the best guesses would be those that are the most widely shared. High-frequency items in the lingua franca are good candidates: they have the best chances of being known to both. Indeed, a distinct preference for the most frequent vocabulary has been attested in ELF.

A similar view is expressed by Hall (2017: 79), who claims that ELF users "engage successfully in joint cognition because of shared communication strategies, a collaborative disposition, and the deployment of linguistic resources shaped by similar Englishing experiences (possibly in the form of overlapping sets of abstract rules distilled from these experiences)".

Such sets of abstract rules can be related to the virtual language (see Section 1.2.1.2.) if the latter is rephrased from a usage-based perspective. As Hall notes (2017: 76–77),

> there are indications that Widdowson and Seidlhofer conceptualize it [i.e. the virtual language] as mentally constituted and indeed consistent with UBL: Widdowson contends that '['ELF users'] are performing on the basis of their knowledge/awareness of virtual rules which, as learners, they have somehow abstracted out of the actual language data they have been taught' (2010, personal communication; cf. also Seidlhofer, 2011: 120). In this view, English is understood not as the 'conventional units' shared by its users, but rather the abstract rules they construct developmentally and employ to formulate utterances.

In other words, the virtual language should be perceived as abstract rules, or rather schemas[16], which hold together

> a mental repertoire of possibilities for novel English constructions determined 'bottom-up' by individual experience. On this interpretation, there will be as many 'virtual Englishes' as there are users of English, the degree of variation between them constrained by (degrees of) mutual intelligibility, influence from other language knowledge (similects), and (conscious or unconscious) sensitivity to conventional norms (Hall 2017: 80).

Concurrently, yet another constraint is put forth by Pitzl (2018), who shifts the focus from linguistic to embodied experience. Her proposal builds upon the works of Lakoff and Johnson (1980: 117), who concentrate on "basic domains of experience" structuring human thought and thus influencing the way humans talk about the world. Among such experiences, Lakoff and Johnson (1980) enumerate, for instance, those gained by means of "our bodies (perceptual and motor apparatus, mental capacities (…)) [or] our interactions with our physical environment (moving, manipulating objects (…))". Available to all people (see Section 3.1.2.2.1.), irrespective of their linguistic and cultural background, embodied experiences are "potentially (…) relevant as a shared and common resource in ELF" (Pitzl 2018: 57).

Together with the acknowledgement that relatively stable units are shared by ELF users, a focus on emerging ELF proposes that ELF communities should be perceived as communities of practice[17], i.e. "groups of individuals who interact with each other through a shared communicative repertoire to achieve a common goal" (Ehrenreich 2018: 37). A shared repertoire in ELF is mainly associated with groups whose interactions are perceived as relatively stable, such as families or academic communities (see Section 2.1.1.) who do not have a language in common (Mauranen 2017). As Mauranen (2021: 76) suggests, "[t]he negotiability of norms can be captured in observing interaction in educational environments, as (...) research fields seem to be determining norms of writing, but not necessarily with reference to Standard English".

In fact, drawing comparisons between ELF and (Standard) English raises substantial criticism among ELF scholars. For instance, it is believed that if ELF is predominantly learnt in the classroom, it should not be compared to a native language

[16] Concurrently, Hall (2017) suggests that such abstract rules are similar to schematic constructions in a Cognitive (Construction) Grammar approach (see Section 3.1.2.2.2.). As Hall claims (2017: 80), "[a]t the abstract end of the continuum, constructions resemble rules, in the sense of regularities that users extract from lexical material". The constructionist approach and constructions are discussed in Chapter 2.

[17] The traditional notion of a speech community is perceived as irrelevant in the context of a global lingua franca since a speech community "is largely local, monolingual, as well as non-mobile" (Mauranen 2017: 10).

which develops through naturally occurring interactions within a speech community. As Alptekin (2011: 159) states, "the 'what' and the 'how' in ELF should not be judged in relation to the 'what' and the 'how' in ENL". Likewise, House (2003: 573) suggests that "the yardstick for measuring ELF speakers' performance should therefore rather be an 'expert in ELF use'".

On a more general note, if ELF is to be liberated from the norms of (Standard) English, i.e. if English is not "to be understood as the overarching framework within which ELF communication takes place", it seems that "the broadly emergentist position on ELF" aligned with a usage-based approach is the only possible choice (Jenkins 2015: 67). In other words, it is only if English is seen "as one among many other languages, one resource among many, available but not necessarily used" (Jenkins 2015: 77), that ELF can develop its full potential as an emergent and emerging complex system.

1.3. Concluding remarks

To sum up, the primary objective of Chapter 1 was to show the complexity behind characterizing and theorizing ELF. Thus, in Section 1.1, ELF was presented as a social phenomenon based in language contact and adequately described along three continua of lingua franca features, with an emphasis on its uniquely international status as a natural lingua franca. Simultaneously, ELF's uniqueness was seen as sustained by the processes of globalization and digitalization, which are not only responsible for the continuing communicative success of ELF but also for its increasing complexity. Moving beyond ELF understood as the result of language contact, Section 1.2. concentrated on a usage-based perspective upon ELF, which, admittedly, duly highlights and adequately captures ELF's inherent complexity and dynamicity. Employing a useful contrast between competence-based and usage-based approaches, ELF, its users and their characteristics were discussed as evolving concepts. For instance, it was demonstrated that ELF could be seen as an interlanguage of a particular learner, a legitimate realization of the virtual language associated with a given user, a repertoire in flux applied among transient international groups or a repertoire used among communities of practice. On a more specific note, ELF was seen as fluctuating between emergent and emerging repertoires of linguistic (and paralinguistic) units recruited from experience and guided by subjective and/or intersubjective cognitive processes.

What clearly transpires from the above summary is a set of continua underpinning ELF theorizing. Already visible in Section 1.1., these axes gain special prominence in

Section 1.2., and lead to the conclusion that ELF should be characterized and researched as both relative and universal, social and cognitive, collective and individual, emergent and emerging. To develop such a balanced overall approach, a number of smaller-scale commitments need to be fulfilled, including, among others, a proper integration of embodiment (see Section 3.1.1.) into ELF theorizing and, relatedly, a motivated synergy between linguistic and paralinguistic units. As evidenced by the discussion so far, such commitments can be detected in ELF theorizing albeit not systematically, which is partly due to the limitations of the general perspective from which ELF has been viewed in Chapter 1. Thus, in order to obtain a more comprehensive picture of ELF, the next chapter focuses on discussing English as a lingua franca within the academic domain, which, as one of the most developed areas in the context of ELF, will allow for a deeper exploration of the general continua and the specific commitments outlined in Chapter 1.

Chapter 2
Academic English as a lingua franca within a usage-based approach

2.0. Outline

The aim of Chapter 2 is to discuss academic English as a lingua franca (henceforth academic ELF) from a usage-based perspective. The chapter is divided into two primary sections, presenting the academic domain and the role of ELF within it (Section 2.1.) and exploring two usage-based approaches currently developed in academic ELF research (Section 2.2.).

More precisely, Section 2.1. offers a discussion of what the academic domain entails and indicates its dependence on ELF. Concurrently, it also highlights how ELF, fostered by the internationalization of higher education, shapes and is shaped by the academic community, ultimately emerging as its shared repertoire, which can be adequately captured by means of a usage-based approach.

Section 2.2., in turn, reinforces the validity of a usage-based perspective upon academic ELF in two interrelated steps. Firstly, the section introduces the theoretical foundations of Linear Unit Grammar and Construction Grammar, which correspond to the general assumptions of emergent and emerging perspectives (see Section 1.2.2.), respectively. Secondly, the section demonstrates how these theoretical foundations are empirically verified, thereby indicating the currently adopted methods for exploring academic ELF.

The chapter concludes with a summary of the key assumptions presented in the above sections, pointing to the necessity of integrating the evolving approaches in researching academic ELF to achieve its fuller understanding.

2.1. Characterizing academic ELF

As indicated in Chapter 1, the academic domain represents one of the spheres which influences and is influenced by ELF. Consequently, to understand the nature of academic ELF, it is necessary to indicate what the academic domain entails and what characteristics of ELF are highlighted when an interplay between ELF and academia is taken into account. These two issues are discussed in Sections 2.1.1. and 2.1.2. below.

2.1.1. The academic domain

To begin with, the contours of the academic domain can be broadly delineated by considering all individuals and activities associated with higher education and/or research endeavors. In other words, academia can be defined as "a community of institutions, the individuals engaged in high level education and research activities and the education and research activities themselves" (Lagoutte and Soskin 2018: 5). Within this broadly delineated community, scholars, for instance, shoulder the responsibility for, among others, conducting and disseminating research through such means as conference presentations or publications.

Importantly, research dissemination involves the collaboration of many individuals fulfilling various roles at various stages. Initially, as indicated by Canagarajah (2018: 44), authors of publications tend to be supported by other scholars who, for instance, provide additional materials or advice, making the final project "a tapestry of many voices". Next, there is the impact of reviewers and editors, who further shape not only the publication's content but also often its layout (Bennett et al. 2020).

Thus, communities engaged in academic activities are perhaps best described as a network of overlapping groups representing the academic domain. As suggested by Mauranen (2010: 7), the academic domain encompasses communities that are "multi-centred, without one hub to revolve around, but with many overlapping hierarchies". Thus, the academic domain can be characterized by communities based on, among others, geographical and disciplinary criteria. In other words, the academic sphere encompasses communities within a specific country or a continent, e.g. scholars in Poland or in Europe, or groups united by shared academic interests, e.g. scholars engaged in such academic disciplines[18] as, for instance, linguistics or literary studies.

However, delineating an academic discipline[19] is not necessarily based on clear-cut criteria. For instance, Becher and Trowler (2006) state that determining whether statistics is sufficiently distinct from the discipline of mathematics to constitute a separate field may depend on the opinions of academic institutions or the recognition of journals dedicated to statistics. Concurrently, Sīle et al. (2021: 66) note that the categorization of journals into specific disciplines involves diverse methods. As they elucidate, attributing a journal to a given discipline may be equivalent to, for example, manually examining its content, e.g. titles, keywords, or abstracts, to identify relevant disciplinary information.

[18] When the word discipline is used, it is seen as equivalent to a disciplinary community.
[19] Disciplines and disciplinary communities are seen as equivalent and thus used interchangeably.

Alternatively, one can take "a more basic and pragmatic approach" (Sīle et al. 2021: 66), which involves reliance on well-known national or international classifications.

In fact, the academic domain has always been characterized by an international factor, i.e. all "academic communities are international by nature" (Björkman 2013: 6–7), and thus need a lingua franca for effective communication. In the past, as signaled in Section 1.1.3, one of the academic lingua francas was Latin, considered the first lingua franca of the academic domain, which dominated until the 17th century (Lindberg 1984, in Björkman 2013). Subsequently, both Latin and French were applied, slowly leaving space for other academic lingua francas such as German or English. By the 19th century, French, English, and German were all employed and the situation changed in the mid-twentieth century (Björkman 2013) since the dependence on English in the world significantly increased (Mauranen 2010). As a consequence, the importance of English in the academic domain also increased and English became the dominant lingua franca in this field (Björkman 2013). In fact, "for the first time in recorded history the entire known world has a shared second language of advanced education" (Coleman 2006: 6).

2.1.2. ELF in the academic domain

As noted by Björkman (2008: 103), "English is now the overwhelmingly dominant language in academia since academic communities use English as the default language". Admittedly, the establishment of English as a dominant lingua franca of academia (Mauranen 2015) is seen as closely related to the internationalization of higher education (Jenkins 2015).

To begin with, universities around the world implement English-medium courses and degree programs and become shared places for people with different native languages, where English is both dominant and co-shaped by the other languages present, i.e. thriving as an academic ELF (Jenkins 2013). Simultaneously, there is "an increasing pressure on academics to disseminate their research internationally in English" (Lorés-Sanz 2016: 54). For instance, Björkman (2013: 10) claims that "[f]or scientists today, publishing findings automatically suggests writing the articles in English since most journals, even in non-English speaking countries, require articles in English". Likewise, Mauranen (2021: 63) states that "individual researchers are under pressure to publish in English, because universities compete with the number and quality of their international publications".

Moreover, according to Hyland (2015: 45), publishing in English is now "more than a choice; it has come to designate research of a high academic quality deemed worthy of a place in globally accessible peer-reviewed journals". In the same vein, Gotti (2021: 146) states that research disseminated in a language other than English is now perceived as "local scholarly products providing only a marginal contribution to the mainstream". Thus, even peripheral journals try to attract at least some international attention by, for instance, publishing abstracts of research articles in English. As Flowerdew (2022: 579) puts it, "[t]hey do this in order to create links between the contents of peripheral journals and the international counterparts published in the Anglophone centre and at the same time connect with readers in other geographical areas and languages in the periphery". Concurrently, since the editors in international journals are not necessarily native speakers of English, the development of academic ELF is fostered[20].

In fact, the idea of non-native speakers' contribution to the development of academic ELF is highlighted by Mauranen (2010: 11), who recognizes international speakers "of high status" and their strong "influence on what is regarded as appropriate academic language", i.e. the fact that influential ELF users "may affect other ELF speakers' notions more than (…) ENL speakers". To illustrate, Mauranen et al. (2016: 48) suggest that

> academically senior people would make comments and instigate corrections on points of language, without consulting native English speakers even if present. This, then, points to changing sources of norms in academic ELF contexts, and raises important questions about who gets to decide what 'good' academic language is like.

In a similar vein, Yilmaz and Römer (2020: 61) claim that the impact of ELF scholars "has been documented to be increasingly visible both in unedited and unpublished research writing (…), published research articles (…), as well as self-published research writing in academic blogs".

Apparently, the native/non-native distinction seems altogether irrelevant in the context of academic ELF since, as argued by, among others, Mauranen et al. (2010), native speakers of academic English do not exist, and both native and non-native writers develop their academic skills in a similar way. In other words, the distinction between

[20] Hence, Flowerdew (2022: 579) remarks that "[i]t would be interesting if research could be extended beyond peripheral journals published in local languages to look at such journals publishing in English, of which there are increasing numbers".

native and non-native users seems to be irrelevant as the idea of experience and, thus, expertise (in ELF) use becomes prominent[21].

Hence, if "the yardstick for measuring ELF speakers' performance (…) [is] an 'expert in ELF use'" (House 2003: 573) (see Section 1.2.2.2. for details), then a usage-based approach becomes particularly fitting. In other words, if academic ELF is defined as a repertoire of linguistic resources co-constructed and shared by members of a heterogeneous and multi-layered community, perhaps best referred to as imagined communities (Mauranen 2017)[22], then it can be felicitously researched as emergent and/or emerging form-meaning pairings.

2.2. Researching academic ELF within a usage-based approach

As described in Section 1.2.2., a usage-based approach is in fact a family of approaches sharing a common set of assumptions entailing, among others, the inseparability of language knowledge and language use, the role of domain-general cognitive processes in guiding the emergence of linguistic units from language use, and the centrality of form-meaning pairings for capturing emergent and/or emerging patterns at individual and/or collective levels.

However, while currently considered "mainstream" (Yilmaz 2020: 3), a usage-based approach is not symmetrically employed for theorizing academic ELF, with some models only recently (re-)gaining the recognition they deserve. Arguably, both Linear Unit Grammar (Mauranen and Sinclair 2006; Mauranen 2016, 2019) and Construction Grammar (Goldberg 1995, 2006, 2019) provide affordances which are of major importance for a systematic description of academic ELF, particularly with reference to the emergent and emerging perspectives (see Section 1.2.2.) but whose full potential has not yet been adequately exploited. To be more specific, with reference to Linear Unit

[21] The irrelevance of the native/non-native distinction is also supported by the fact that, as asserted by Lorés-Sanz (2016: 54–55), "all writing in international journals is, by definition, for an international rather than an ENL (English as a native language) audience". However, even though "the vast majority of readers and writers are not native speakers of English" (Mauranen et al. 2010: 184), many journals still insist on having texts proofread by native speakers prior to publication (Ammon 2000; Hu 2004; Li and Flowerdew 2007). As a consequence, Ingvarsdóttir and Arnbjörnsdóttir (2013: 123) claim that "ENL writing standards [still] seem to be the only accepted norm". On the other hand, though, there appear journals, such as the *Journal of English as a Lingua Franca*, which "expect authors to submit manuscripts written in an English which is intelligible to a wide international academic audience, but it need not conform to native English norms" (Flowerdew 2022: 578). All in all, then, as Flowedew (2022: 580) concludes, "by hook or by crook" ELF is "likely to increasingly appear in the academic domain".

[22] Mauranen (2010: 7) explains that imagined communities are those whose "members do not all know or meet each other personally, but people nevertheless see themselves as members of such groups, and identify with them".

Grammar (henceforth also LUG), Mauranen (2016: 96) admits that the model "has not been [generally] developed or tested by other scholars very much yet", while Yilmaz (2020: 19) observes that theorizing academic ELF within the framework of Construction Grammar (henceforth also CxG) "is still in its infancy". Thus, Section 2.2.1. presents basic theoretical underpinnings of the two usage-based grammars with a view to uncovering paths along which their further development along the emergent-emerging cline could proceed. Section 2.2.2., in turn, enriches the theoretical scaffolding with empirical details. To be more specific, the methodological identity of a usage-based approach to academic ELF is described in terms of data-gathering strategies and ways of analyzing empirical data.

2.2.1. Theorizing academic ELF along the emergent-emerging continuum

As Mauranen and Sinclair (2006: 31) observe, " [s]ome features of Linear Unit Grammar have been noted as similar to those of Construction Grammar", including "[t]he holistic approach, the lack of a predetermined hierarchy of units into which everything has to fit, the separation of relationships external to a construction and those internal to it". Nevertheless, while correspondences between LUG and CxG can indeed be noticed, with a possibility of situating both grammars within a constructionist approach in a broad sense (Habjan 2016; Ungerer 2023), the two models begin their respective theorizing at the opposite ends of the emergent-emerging continuum and thus, though potentially reconcilable, are kept apart here for the sake of clarity of exposition and discussed in Sections 2.2.1.1. and 2.2.1.2., respectively.

2.2.1.1. Linear Unit Grammar (LUG)

Linear Unit Grammar, though focused on spoken English, is in fact "intended to apply to all varieties of English, whether written or spoken, whether standard or non-standard, whether specialised or general" (Mauranen and Sinclair 2006: 15). To be more specific, it is a usage-based model based on "a single supposition—that chunking is a natural and unavoidable way of perceiving language text as it is encountered" (Mauranen and Sinclair 2006: 6). Thus, LUG takes an emergent approach to linguistic units and argues that "a real-time perspective is vital for developing novel and more realistic models of language processing", i.e. for the "need to incorporate the online experience of language into accounts of its meaning and structure" (Mauranen 2018: 26).

LUG is essentially a three-stage approach. As Mauranen (2018: 26) explains, the first phase involves "the intuitive, spontaneous segmentation of ongoing speech" or writing, and is "independent of an analyst's view". The second one, in turn, "imposes an analytical, linguistic model on the chunks that the non-analytical first phase has produced. This is the analyst's perspective". The third stage synthesizes "the outcomes of the earlier steps. It also helps serve as a bridge between LUG and other grammars. Moreover, it also narrows the gap between descriptions of spoken and written text" (Mauranen 2016: 93–94).

The key cognitive process in LUG, most prominent in the first phase, is chunking, i.e. segmenting incoming linguistic information, often against

> the limited resources of working memory, which Christiansen & Chater (2016) identified as the 'now-or-never bottleneck' of processing: if linguistic information is not processed fast, that information is lost for good. What they suggest is that listeners (as well as speakers) engage in 'Chunk-and-Pass'-processing, and incrementally chunk up speech as fast as possible, using all available information to process current input before new information arrives. For this, hearers make use of a wide array of information, such as different levels of linguistic structure, situational information, and prior knowledge (Mauranen 2018: 27).

At the same time, Christiansen and Chater (2016: 26) emphasize that "Chunk-and-Pass processing is [not] exclusively driven by memory constraints and obeys the rules of the language system". In fact, recent evidence suggests "that chunking is driven by a need for global coherence manifested as Gestalt-like structures, which in turn underlie memory organisation and mirror real-world phenomena". In the same vein, Mauranen and Sinclair (2006: 31) observe that "at the very outset of the interpretative [i.e. chunking] process some hierarchy is imposed on the linearity of text, with a strong tendency to prioritise gestalt patterns where relevant". In other words, at least some interplay between Gestalt perception (see Section 3.1.1.1.) and chunking can be assumed.

In fact, the intuitive process of chunk delineation is facilitated by a number of factors. More specifically, it may involve such factors as

> grammatical, phonological, semantic or cognitive – coming together to bring about a sense of the 'natural' chunk boundary that we seem intuitively to respond to. (…) Plausible linguistic sources are phonological, tonal and intonational patterning, pausing, syntactic boundaries, phrases, breathing patterns, punctuation marks, etc. Undoubtedly paralinguistic and non-linguistic situational sources play a role, too (Mauranen 2016: 83).

For instance, "the reader intuitively gives boundary status to the wider gaps between characters" (Mauranen and Sinclair 2006: 10). All in all, however, the product of chunking, i.e. a chunk, "is a pre-theoretical term, and defining it precisely is not even attempted in LUG or at present" (Mauranen 2016: 82). Still, its feasibility can be

supported by, for instance, results of inter-rater reliability tests of assigning boundaries (Vetchinnikova et al. 2022).

Once chunks are delineated, i.e. "provisional unit boundaries" (Mauranen 2016: 83) are assigned, the second, i.e. analytical, phase begins. This stage involves dividing the chunks into message-oriented (labeled as M) and organization-oriented (labeled as O). The former are concerned with the content of a given communicative encounter, e.g. "you have to use the language", and the latter with managing communication, e.g. "right". M and O elements are then described in more detail as, for instance, complete chunks, e.g. "it's law now", and incomplete ones, e.g. "you have to use their".

The final, or synthetic, stage "results in linear units of meaning (…) [which] come fairly close to units generally recognised in grammars, such as clauses and phrases" (Mauranen 2016: 94). Hence,

> they incorporate the rudiments of hierarchy in that they recognise two kinds of relationship, one of which can constitute a component in the other: the endocentric forms 'textual objects', single entities like noun phrases that must combine with something else to set up a communicative act (streets with shops). The exocentric relationship forms 'textual incidents' by relating two separate entities such as subjects and predicates (there are streets with shops) (Mauranen 2016: 94).

Thus, while linearity is, on the whole, advocated in LUG, minimal hierarchy is also in place.

All in all, LUG is an emergent perspective capturing the online and fleeting nature of linguistic experience. Thus, as Mauranen (2016: 83) puts it, although a number of researchers have applied the term chunk in relation to, among others, "fixed expressions, formulaic sequences, multi-word units, constructions, etc., [t]hese can all be seen as products of the chunking process. What we are concerned with in LUG is not the product in the first instance, but the process". In other words, while "it is indeed possible to identify certain chunks as performing similar functions in different texts" (Mauranen 2016: 83), this is not the main object of interest in LUG. A reversed hierarchy of priorities characterizes Construction Grammar.

2.2.1.2. Construction Grammar (CxG)

Construction Grammar is intended to provide a comprehensive description of grammar, with emphasis "laid on the breadth of coverage" (Maranen and Sinclair 2006: 31) and a "potential for serving as a universal theory" (Östman and Fried 2005: 8). Given this commitment, "most of the research in Construction Grammar has addressed more general

constructions found at the level of a given language" (Hart and Queralto 2021: 534), and possibly shared across other languages, rather than constructions specific to a particular variety or genre (see Section 4.1). Confronted with such an overwhelming task, CxG has, expectedly, grown into a set of related approaches which are, nevertheless, held together by one basic assumption – that "[a] grammar contains nothing but constructions, so the only question is precisely what counts as" a construction (Holmes and Hudson 2005: 270). This question, however, has been and still is answered in a number of ways (see Haspelmath 2023 for an overview), which can be neatly illustrated on the basis of Goldberg's renditions of the notion (for a related discussion see Section 3.1.2.).

To begin with, one of the early definitions proposed by Goldberg assumes that a construction is "a form-meaning pair <Fi, Si> such that some aspect of Fi or some aspect of Si is not strictly predictable from C's component parts or from other previously established constructions" (Goldberg 1995: 4). This early understanding is then modified in the following way: "[a]ny linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency" (Goldberg 2006: 5). Finally, there is the most recent definition of constructions which "are understood to be emergent clusters of lossy memory traces that are aligned within our high- (hyper!) dimensional conceptual space on the basis of shared form, function, and contextual dimensions" (Goldberg 2019: 7).

Clearly, then, the concept of a construction has become more and more inclusive and flexible, moving from "categorical criteria of non-predictability or sufficient frequency" towards a gradient view, i.e. a possibility "that constructions arise through a continuous process that presumably leads to increasingly higher levels of constructionhood rather than to a sharp division between constructions and 'non-constructions'" (Ungerer 2023: 6). Accordingly, even a single exemplar, a.k.a. a construct, i.e. "a transient neural event, which occurs and then is gone", is likely "to leave a trace of its occurrence" which can be entrenched into a construction (Langacker 2017c: 54). Ultimately, then, CxG should "distinguish entrenched constructions that are stored in the long-term memory from form-meaning pairings that are assembled in the working memory (online constructions). Once this distinction is made, the precise role of (…) constructs and the nature of (…) constructions can finally be disentangled" (Hoffmann 2016: 1).

The key cognitive process in CxG is "the elusive concept of entrenchment" (Divjak 2019: 137), a.k.a. automatization, unitization, or chunking. Generally understood as the process of strengthening memory representations, entrenchment is, on the whole, facilitated by frequency of repetition (see Section 3.1.2.2.1.). Simply put, "the more frequently a given word or structure has been processed, the more deeply the patterns of associations that are activated for processing will become entrenched. Therefore, frequency of rehearsal and repetition of whatever we can identify as being similar in some respect (...) are expected to contribute to the learning and consolidation processes" (Schmid 2020: 216). Still, "[u]nder some conditions a unit (...) can be learned from a single exposure. Thus the sheer number of usage events may be less important than some measure of cumulative psychological impact" (Langacker 2008: 220). Moreover, salience and embodiment[23] effects influence entrenchment, with "basic principles of perception and attention, e.g. figure-ground segregation and other Gestalt principles" (Schmid 2020: 222) clearly facilitating the process.

Given the complex nature of entrenchment, it is important for "researchers to be explicit about whatever strategies they use to mark off a category of 'constructions'" (Ungerer 2023: 10). A crucial question that thus needs to be answered concerns the respective roles of top-down and bottom-up processes. To be more specific, analysts should clearly distinguish between putative, or assumed, constructions and those that are purely data-driven in order to adequately describe an interplay between already-emerged and newly-emerging patterns (see Section 1.2.2.). A related issue is whether "suitable thresholds have to be determined in a top-down fashion based on the analyst's intuition, or can they be inferred in a bottom-up way from the properties of the linguistic units themselves?". Finally, is it "possible to determine a 'universal' threshold for constructionhood that applies across different phenomena, or are thresholds necessarily defined in relation to a given analysis (and its specific objectives)?" (Ungerer 2023: 10).

Once a construction is identified, or rather situated along the continuum of constructionhood, it can be further described as, for instance, more specific or schematic or more simple or complex (see Section 3.1.2.2.). As a result, constructions range from, among others, specific (a dog) and schematic phrases (DET NOUN), or from simple words to complex discourse genre constructions (see Chapter 4), and participate in a

[23] For a definition of embodiment see Section 3.1.1.1.

number of relations, e.g. part-whole and schema-instantiation (see Diessel 2019 for an overview), creating a complex hierarchical architecture, i.e. a constructicon.

All in all, CxG is an emerging perspective capturing the offline and stabilizing nature of linguistic experience. While its main interest is undoubtedly in types, or constructions, whose role is primarily to sanction tokens, or constructs, there is definitely a place for online constructions in the model.

2.2.2. Exploring academic ELF along the emergent-emerging continuum

As noted in Section 1.2.2. a usage-based approach places actual language use in research focus. In other words, "[u]sage-based linguistics attributes a central position to usage events as the empirical foundation of language research and linguistic theory" (Tummers et al. 2005: 229). Thus, the theoretical scaffoldings of LUG and CxG are now filled with empirical details, i.e. strategies of data-gathering (Section 2.2.2.1.) and ways of data analysis (Section 2.2.2.2.), in order to both reinforce the shared usage-based identity of the two grammars and accentuate their particular methodological choices.

2.2.2.1. Collecting academic ELF data

According to Mauranen et al. (2010: 184), "[c]oming to grips with academic lingua franca English requires a good database" since a large corpus, while still capturing variation, can unveil preferences emerging among ELF users. As a result, a number of academic ELF corpora have been created and are now widely utilized in order to describe both emergent and emerging units within spoken and written academic ELF.

While academic ELF corpora can indeed be neatly divided into spoken or written, such a separation is not consistently adopted. Thus, on the one hand, Ranta (2009) advocates separating data pertaining to writing and speaking since the latter is associated with language use generated spontaneously in real-time without the opportunity for editing, while the former enables pauses for reflection and hence language use undergoes substantial editing. On the other hand, Mauranen (2013: 30–31) argues that speaking and writing exist along a continuum and allow for at least some overlap[24]. In particular, given the role of digitalization, the boundary between speaking and writing is being blurred.

[24] A similar approach beyond ELF research is developed by Bohát (2018: 60), one of the compilers of the COHAT corpus, who admits that there are traditional written data, written to be spoken data (e.g. speeches, presentations) and spoken data.

Research blogs, for instance, involve at least some "spontaneity of spoken language". On a more general note, Kuteeva (2023) claims that "academic discourse has (…) evolved to incorporate interactional or dialogic features into the writing counterparts"[25].

In what follows, two academic ELF corpora exploited by, among others, scholars working within LUG and CxG frameworks, i.e. ELFA (a corpus of spoken academic English as a lingua franca) and WrELFA (a corpus of written academic English as a lingua franca) are presented. While the two databases are discussed separately, primarily for the sake of clarity, it is also assumed, given the fuzzy boundary between speaking and writing, that particular solutions adopted for the compilation of one type of corpus can be taken as largely applicable in the case of the other.

In fact, as noted by Pitzl (2018), the creation of ELF corpora necessitates clear compilation principles[26]. Such principles, in accordance with the general assumptions of corpus compilation (see Lüdeling and Kytö 2008), concern, among others, the delineation of a sampling unit, i.e. the type of data included in a corpus (Egbert et al. 2022) or the choice of criteria, i.e. internal or external[27], determining the selection of a sampling unit (Breiteneder et al. 2006: 164). At the same time, the principles may determine how corpus balance[28], i.e. "the [relatively equal] proportions of the different samples" (Ädel 2020: 5) is achieved by corpus compilers, or what data descriptions, i.e. additional information about tokens (Evans 2018), are incorporated into a corpus. Since the majority of available academic ELF corpora are "disposable" (Bernardini and Baroni 2004), i.e. "quickly constructed for a specific purpose and rapidly discarded" (Hunston 2008: 154), and thus lacking compilation principles, ELFA (Section 2.2.2.1.1.) and WrELFA (Section 2.2.2.1.2.), which come with clearly delineated criteria, are particularly relevant when discussing data-gathering strategies characteristic of a usage-based approach to academic ELF.

[25] In fact, given the role of digitalization, the way in which research dissemination is achieved develops. Hence, for instance, graphical (Guinda 2022) abstracts have recently been established.

[26] Simultaneously, given the need for new ELF corpora, Laitinen et al. (2019) stress that the principles of corpus compilation should be described as precisely as possible because fine-grained descriptions of all the decisions made facilitate the replicability of procedures across, for instance, various countries.

[27] Internal criteria are typically associated with the linguistic features of a sampling unit, such as the occurrence of prepositions, while external criteria involve the non-linguistic features of a sampling unit, such as its socio-cultural context (Breiteneder et al. 2006).

[28] However, as Ädel (2020: 5) comments, the notion of, among others, balance is "scalar and vague (…) so there are no hard and fast rules for achieving (…) balance in a corpus". Therefore, Kegalj and Borucinsky (2021: 219) claim that "[c]arefully defined criteria for selecting texts that are to be included in the corpus are important to ultimately design a corpus that is (…) balanced as much as possible, and that satisfies the needs of the study".

2.2.2.1.1. The ELFA corpus and spoken academic ELF data

The ELFA corpus stands for the corpus of English as a Lingua Franca in Academic Settings, i.e. "the first corpus of academic English spoken as a lingua franca" (Mauranen et al. 2020: 12) which is the "most widely studied corpus in the field" (Mauranen et al. 2016: 46). The ELFA corpus includes data (i.e. 1 million words) representing academic interactions among native and non-native users of English at four Finnish universities, i.e. Helsinki, Tampere, Aalto, and Tampere Technological University (Mauranen 2015). Its sampling unit is a speech event type, i.e. a term roughly equivalent to discourse genres[29], which, as Mauranen (2006) claims, allows for taking into consideration types of academic speech that are both "much further established", e.g. lectures, and less common or less studied, e.g. workshops or panels.

In total, the corpus includes 9 academic speech event types, i.e. seminar discussions, PhD thesis defense discussions, lectures, conference presentations, seminar presentations, conference discussions, lecture discussions, PhD thesis defense presentations, and panel discussions, with seminar discussions (33%) and PhD thesis defense discussions (20%) constituting the largest proportion of tokens. The recognition of the aforementioned academic speech event types was based on external criteria and involved "the distinctions and labels that the university community uses of its own discourses and genres" (Mauranen 2006: 151). In other words, if a particular academic community identified a given speech event type as a lecture, it was taken into account and described in the corpus as such. At the same time, the ELFA corpus specifies that the choice of a given speech event type was motivated by three factors, i.e. prototypicality, influence, and prestige. The first one refers to the extent to which a given speech event type is shared across various academic disciplines (e.g. lectures which are well-known for various disciplines). The second one is seen as reflecting the number of participants associated with a given speech event type (e.g. panel discussions characterized by a great number of participants). The last one refers to the high status of a given speech event type among academic communities (e.g. conference discussions). Moreover, the authors state that they aimed to include data from various academic disciplinary domains, e.g. Art or Technology, which can be further divided into disciplines, e.g. electrical engineering, and sub-disciplines, e.g. organic chemistry (Mauranen 2006), the choice of which was

[29] Due to their inclusive nature, i.e. encompassing well-known (emerged) and new (emerging) genres, speech event types can be taken as equivalent to discourse genre constructions introduced in Chapter 4.

influenced partly by practical considerations, i.e. their availability in the relevant universities.

Consequently, given the inclusion of a disciplinary division in the corpus, the authors of ELFA claim that corpus balance can be determined on the basis of, among others, different disciplinary domains. For instance, the corpus description on the website indicates that the ELFA corpus contains data pertaining to, among others, 29% of Social Sciences, 19% of Technology, or 17% of Humanities, which "give a clear picture of the overall balance of the corpus" (Mauranen 2006: 152).

Moreover, each token in the corpus is described by means of two types of data. Firstly, the corpus provides information about gender, age, academic role or the native language of participants at the beginning of each token. Secondly, each token contains supplementary information provided in the file to indicate, among others, the beginning and the end of utterances as well as pauses, laughter, hesitation or "even spelling out a word or acronym etc., as letters"[30].

2.2.2.1.2. The WrELFA corpus and written academic ELF data

When it comes to WrELFA, i.e. "the only written ELF corpus" (Laitinen 2018: 61)[31], it contains data produced by both native and non-native users of English (Rowley-Jolivet 2017). To be more specific, WrELFA encompasses 1.5 million words pertaining to three academic genres, i.e. research blogs, PhD examiner reports, and unedited research articles, each of which constitutes a separate sub-corpus.

The authors of the corpus indicate that the compilation of the WrELFA corpus followed the principles of the ELFA corpus (http://surl.li/vehzbo). Concurrently, for each sub-corpus, there are separately discussed external criteria. For instance, the corpus compilers indicate that they were interested in collecting only the tokens of research blogs whose authors were non-native users of English and whose authors were not professionally associated with a country in which English is spoken as a mother tongue[32]. Tokens of PhD examiner reports, in turn, were taken into account irrespective of the native language of participants. In other words, the corpus also includes tokens of reports

[30] This information is provided in the ELFA transcription guide (2004).

[31] In fact, as noted by Ranta (2017: 251), there is a noticeable lack of written ELF databases and the compilation of WrELFA was seen as the first step towards providing ELF corpora devoted to written ELF in general. Against this background, Laitinen (2018) states that "[t]o complement these first-generation corpora [i.e. spoken ELF corpora and WrELFA] and to compensate for the fact that 'genuine ELF written text databases are still missing' (Mauranen et al. 2015: 402)", FIN-CE and SWE-CE are currently compiled.

[32] The details about the authors are gathered through platforms such as LinkedIn or private correspondence.

written by native English users, which were created in "a context where English [was] the lingua franca between a Finnish university and academics from 33 identified L1 backgrounds". Finally, tokens of unedited research articles were incorporated into the corpus if they were written by non-native users and if they were not checked or corrected by a native speaker.

Furthermore, only tokens which were freely accessible online or whose authors had given consent were incorporated into the corpus. More specifically, because they were publicly available, tokens of research blogs were obtained through the researchblogging.org platform without consultations with their authors. In contrast, when it comes to tokens of PhD examiner reports and unedited research articles, which were not freely accessible, their authors were asked for permission concerning their publication.

As WrELFA utilizes a broad binary classification of tokens into disciplines in the sciences ("Sci") and disciplines in social sciences and humanities ("SSH"), the overall balance of the corpus is presented with reference to the above-mentioned division, indicating that the corpus is reasonably well-balanced as the Sci category represents 55% and SSH 45% of the data. Moreover, information concerning balance is provided with reference to the sub-corpora. For instance, while the sub-corpus of PhD examiner reports is fairly well-balanced, the subcorpus of research blogs is heavily skewed towards the category of Sci and the sub-corpus of unedited research articles is skewed towards SSH (Rowley-Jolivet 2017). As the authors indicate, they "attempted to compile a balanced sample of papers between the sciences (…) and the social sciences and humanities (…). However, the texts categorised as SSH were found to be much longer on average than those labelled Sci" (http://surl.li/yokpft).

When it comes to the description of tokens, additional information includes data about active hyperlinks in blogs or the use of text formatting, e.g. bold or italics (http://surl.li/vehzbo). Concurrently, there is a separate document describing the latest addition to WrELFA, i.e. the sub-corpus of unedited research articles, in which various types of data incorporated into the corpus are discussed (Carey 2015), including the native language of the author or the role he or she plays in the academic domain, e.g. a research student. At the same time, the sub-corpus provides data concerning the discipline to which a given token is attributed, the year in which it was created, or the number of its authors. Moreover, the corpus provides information concerning, among others, the use of foreign words or block quotes in a given token.

2.2.2.2. Analyzing academic ELF: between emergent and emerging perspectives

While academic ELF corpora indeed provide large databases in which form-meaning pairings produced by members of a heterogeneous and multi-layered community, i.e. an imagined international academic community, can be traced, these naturally-occurring resources can be analyzed from either an emergent or an emerging perspective. In what follows, the former approach is illustrated by means of analyses of (predominantly) spoken academic ELF data assembled in ELFA and conducted within the framework of LUG (Section 2.2.2.2.1.) and the latter through studies on written academic ELF data gathered in WrELFA and based in CxG (Section 2.2.2.2.2.). To consistently reveal the research designs of the chosen analyses, the discussion encompasses their aims, procedures, results and conclusions, thus providing detailed descriptions complementing the theoretical foundations set in Sections 2.2.1.1. and 2.2.1.2., respectively.

2.2.2.2.1. LUG-based studies of spoken academic ELF data

Apparently, although LUG is intended to apply to both written and spoken data, the latter appears more explored than the former. At the same time, while the main focus of the LUG model is the online experience of language and the cognitive process of real-time chunking, empirical research seems to indicate that LUG can be also felicitously applied to analyzing the offline mode of cognition and the product of the chunking process, e.g. a multi-word unit. Both tendencies transpire in the four studies discussed below.

The first study illustrating how the theoretical foundations of LUG can be empirically tested and expanded is the analysis conducted by Vetchinnikova et al. (2022). The study builds on the central premise of LUG, i.e. the idea that when humans process linguistic information they face the challenge of making sense of the input while simultaneously receiving new material. Consequently, there is a need for rapid processing to avoid information loss, which is achieved by the inherent capacity of humans to divide sensory input into smaller chunks. Hence, as the authors emphasize, it is crucial to analyze the real-time segmentation of incoming aural signals into linear, or temporal, groups. Following Sinclair and Mauranen's (2006) suggestion "that anyone fluent in the language would chunk it up in approximately the same way" and realizing the limitations of LUG resulting from its reliance "on only two coders" (Vetchinnikova et al. 2022: 4), the authors aim to empirically validate the hypothesis that chunk boundary identification is largely convergent. To be more specific, Vetchinnikova et al. (2022) attempt to determine the level of agreement in perceptual chunking among non-native English users whose task is

to intuitively indicate boundaries of chunks while listening to different fragments of authentic academic speech.

The study relies on extracts from three corpora, i.e. ELFA (see Section 2.2.2.1.1.) VOICE[33] and MICASE[34], which together "reflect the kind of English people are typically exposed to in today's highly international universities" (Vetchinnikova et al. 2022: 6). Three datasets, i.e. A, B and C, in the form of transcripts and corresponding audio clips, are selected from the corpora. To additionally investigate how the quality of speech influences agreement on chunk boundaries, different selection procedures are employed for sets A and B, and for set C. In the case of the former, explicit criteria are followed to gather a homogeneous collection of clear, fluent, and easily understandable extracts. These criteria include: avoiding unintelligible or unfinished words, laughter, long pauses, overlapping speech, speaker changes, frequent hesitations, repetitions, and specialized or low-frequency vocabulary. In contrast, in the case of the latter dataset, no particular selection procedures are devised and instead "a random sample of extracts representative of the corpus and the original audio clips" (Vetchinnikova et al. 2022: 7) is provided.

To gather chunking data, a specialized tablet application called ChunkitApp is developed. This application synchronizes transcripts of audio clips with the audio version. Hence, the task of the participants (149 non-native English users from a variety of L1 contexts) is to listen to the audio version and simultaneously follow the transcript to mark chunk boundaries between words in real time by clicking "~" symbols (see Figure 1). Each participant's chunking task results is a sequence of words with boundaries marked as 1 and unmarked as 0, e.g. "I (0) mean (1) there's (0) a (0) school (0) down (1) not (0) very (0) far (0) from (0) here".

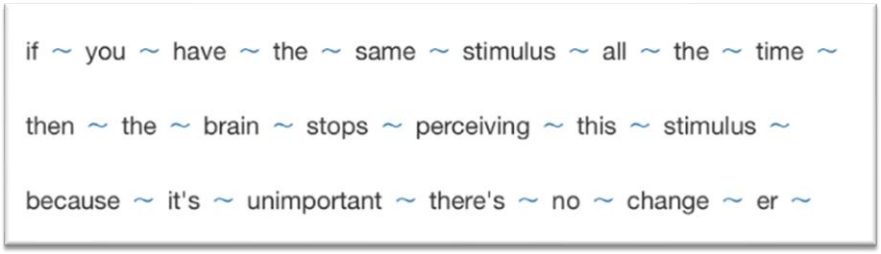


Figure 1. ChunkitApp - user interface (Vetchinnikova et al. 2022: 5)

[33] VOICE, i.e. Vienna-Oxford International Corpus of English, is a corpus of naturally-occuring spoken English as a lingua franca.
[34] MICASE, i.e. the Michigan Corpus of Academic Spoken English, is a corpus of academic speech of native and non-native users of English.

There can be a mark between any two words, so the number of potential boundaries in each transcript is the number of words minus one as no boundary can be added at the end of the transcript. The boundary frequency is how many participants marked a boundary at a specific spot. For example, if 15 out of 45 participants marked a boundary, the boundary frequency is 15, which is further analyzed for agreement between participants.

The results of the study indicate that ELF users tend to agree on chunk boundaries and it is established that the typical chunk length perceived by an individual ELF participant averages between 9 to 10 words. More specifically, in all three datasets, there is a high level of agreement observed in boundary markings among participants, i.e. around 90%. At the same time, a statistical comparison of the level of agreement in three datasets reveals significant differences, i.e. the agreement among participants in sets A and B is greater than in the set C. Hence, the authors conclude that the quality of the extracts influences the agreement among listeners regarding chunk boundaries. In other words, when exposed to clear, easily comprehensible extracts, ELF users tend to align more closely on the delineation of chunk boundaries. Moreover, since the chunks identified by ELF users do not recur in the sets, i.e. users delineate different chunks in the three sets, the authors emphasize the need for distinguishing between online, or perceptual chunking and usage-based chunking, i.e. learning. As Vetchinnikova et al. (2022: 13) state,

> [t]hrough usage-based chunking listeners pick up statistical regularities from the input and in effect learn the language, which can be thought of as an array of such statistical regularities. Perceptual chunking, instead, carves up the input into manageable bits for further processing. These elements do not generally correspond to recurrent multi-word units. If they did, memory constraints would have to be re-thought, since a multi-word unit constitutes one unit by definition, while our memory can process around four according to recent accounts. At the same time, it is likely that learned usage-based chunks play a role in perceptual chunking: at the very least it is clear that a perceptual chunk boundary cannot lie within a multi-word unit, but a more intricate interaction between them is also possible. Given that perceptual chunks do not recur, it is unlikely that they develop into multi-word units with time.

The nature of perceptual chunking and its relation to usage-based chunking is further[35] explored by Vetchinnikova et al. (2023), who essentially postulate "two different chunking processes operating at the same time: on the one hand we draw on the inventory of chunks available to us (usage-based chunking), on the other we segment incoming stream into temporal groups (perceptual chunking)". Focusing on the emergent side of chunking, Vetchinnikova et al. (2023) concentrate on cues, e.g. prosody, syntax, or

surprisal effect, that facilitate human perception of chunk boundaries. More specifically, Vetchinnikova et al. (2023) aim to investigate the extent to which boundaries suggested by specific intonation, clauses, or typical word associations[36] align with the boundaries identified by study participants.

The study relies on the same corpora, i.e. ELFA, VOICE, and MICASE, and involves the use of Chunkitapp application. The task of the participants is to listen to the audio clips while simultaneously marking chunk boundaries on the transcripts displayed on the screen (see Figure 2). The concept of a chunk is left undefined, prompting participants to rely on their intuition. Each audio clip is played only once.

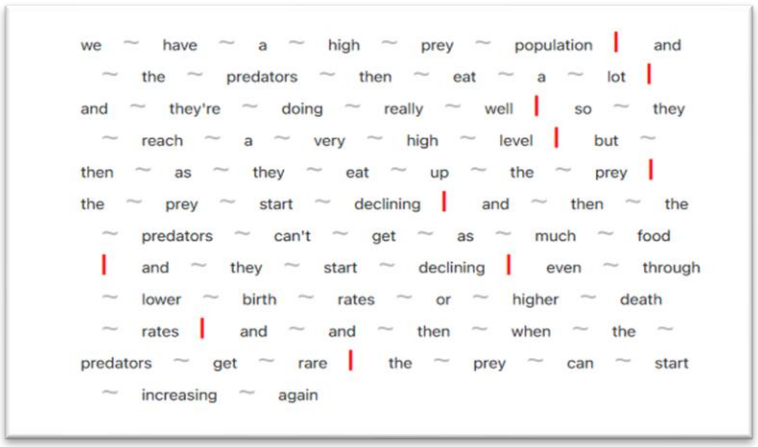


Figure 2. Boundary perception (Vetchinnikova et al. 2023: 7)

The analysis shows that each of the analyzed cues aids in boundary delineation, however, prosody appears to play the most significant role while surprisal effect the least. Concurrently, with reference to surprisal effect, Vetchinnikova et al. (2023: 22) note that

> [t]he results for the effect of surprisal support the proposed distinction between perceptual and usage-based chunking. If perceptual chunks were learned multi-word units, the words before the boundary should be less surprising because they belong to the ongoing unit, and the words after the boundary should be more surprising since they start a new unit. The models show the opposite results: chunk boundaries associate with higher closing surprisal and lower opening surprisal. However, both effects were very small as well as different in direction across extracts. Thus, the relationship between statistical information and perceptual chunk boundaries requires further research.

In fact, a tentative route along which research into online and offline chunking could co-evolve is outlined by Mauranen, who analyzes "English spoken as a lingua franca from two complementary perspectives: linearity and recurrent patterning" (2009:

[36] As Vetchinnikova et al. (2023: 6) explain, "sort is almost always followed by of, so of will be expected by the listener and its surprisal will be low".

230). To be more specific, the study first establishes emergent chunks in one extract from the ELFA corpus and then determines their frequency. As Mauranen (2009: 224) explains, "[t]he extract discussed (…) comes from a seminar discussion randomly picked from the corpus and subjected to a linear reading. The corpus as a whole is used for checking the distributions of recurrent expressions".

In the first step, adopting the procedure devised in LUG (see Section 2.2.1.1.), chunk boundaries are delineated and chunks of varying complexity are proposed, e.g. but | that would allow | er | people | more time to do whatever | yeah | but | I | in my point of view | that was really good | because | @yeah@ | yeah (xx) | that | we are an open | an open market nowadays. Subsequently, the chunks are classified as message-oriented (M and its combinations) and organization-oriented (O and its combinations) in the following way: but OT | that would allow M | er OI | people M | more time to do whatever M | yeah OI | but OT | I M| in my point of view OI | that was really good M | because OT | @yeah@ OI | yeah (xx) | that OT | we are an open M | an open market nowadays M. Importantly, each chunk is related to a meaning, or function. For instance, | er OI | indicates hesitation, serving to regulate the interaction by signaling the speaker's intention to continue speaking, | yeah OI | expresses agreement with the preceding speaker, signaling acknowledgment of what has been said, and | in my point of view OI | indicates that the following text is to be interpreted as an opinion. As Mauranen (2009: 227-228) elucidates, "the OI in my point of view (…) can be viewed in formal terms as a structural blend of two expressions, which share a meaning. From the perspective of ongoing discourse, or language in interaction, its function is to preface an utterance, signaling that the speaker's opinion will follow".

In the second step, the focus is on the chunk "in my point of view". To be more specific, Mauranen explores the chunk's frequency in the ELFA corpus and related databases to argue that it is not "an individual's processing error or a fossilized idiolectal feature" but an emerging tendency. In fact, corpus searches indicate that "in my point of view" appears rather infrequently, i.e. 13 times/10,000 words. However, Mauranen observes that even though these instances do not indicate a feature of ELF adopted universally, they are still noteworthy. To strengthen her argument, Mauranen uses other data sources, including VOICE, to check if "in my point of view" also appears there. Ultimately, Mauranen concludes that she found its "traces" based on other data, suggesting that it might be an emerging tendency because "in my point of view" is used in various interactions by different ELF users. This tendency speaks to the fact

> that ELF is not a chaos of idiosyncrasies which arise at random and may work in their contexts but which are not likely to have a bearing on English as a whole because there is no regularity or systematicity. (...) English spoken as a lingua franca is normal use of natural language, and at the same time has its unique features which constitute an important part of the pool of global English (Mauranen 2009: 2).

Building on the LUG framework in general and Mauranen's (2009) analysis in particular, Carey (2013) focuses on multi-word chunks in academic ELF. More specifically, Carey strengthens Mauranen's (2012) argument against ELF users' overall reliance on the open-choice principle, i.e. constructing utterances in a bottom-up manner (Sinclair 1991; Wray 2002), and in favour of the idiom principle, i.e. retrieving "semi-preconstructed chunks of language from memory" (Carey 2013: 209), often through approximation (see Section 1.2.2.2.). Defined as a cognitive process, approximation is not "a question of choosing to diverge from a conventional phraseological chunk, but instead an involuntary byproduct of cognitive constraints such as memory and real-time access to stored chunks" (Mauranen 2012: 42, in Carey 2013: 210). Importantly, then, approximation is "a reflection of frequency effects in a speaker's accumulated experience with a language", i.e. in both input ("what is heard in one's surroundings") and output ("successful use of handy chunks of language that serve a recurring purpose") (Carey 2013: 210).

As Carey (2013: 210) claims, "[i]t is easy to find one-off examples of approximation in ELF data (…), but it is hard to draw any meaningful conclusions about the frequency of variation or the potential for emerging ELF-specific trends when the conventional phrases are low-frequency expressions in themselves", e.g. "now that you mention it" or are repeated by the same user. For instance, the approximated version of the chunk "as a matter of fact", i.e. "as the matter of fact" appears 21 times in the data but is spoken only by two users in two separate events. Therefore, Carey (2013: 211) states it is essential to analyze highly frequent chunks "found among different speakers, in unrelated times, locations, and events" to identify any emerging ELF-specific units.

Thus, the paper analyzes the most frequently occurring three to five-word organization-oriented, or O, chunks (see Section 2.2.1.1. for details), e.g. "in my view", "so to speak", "on the other hand", derived from the (primarily) ELFA and (secondarily) WrELFA (see Section 2.2.2.1.2.) corpora, supplemented by the VOICE corpus and the MICASE corpus for the sake of reference, i.e. to compare the frequencies of a given spoken chunk in a different ELF corpus (VOICE) and in a native corpus (MICASE). In an attempt to identify the most commonly occurring multi-word chunks of three to five

words in the ELFA corpus, an n-gram list is generated using the AntConc corpus analysis freeware (Anthony 2011). Subsequently, to detect potential approximations of these chunks, repeated searches are conducted on words within each chunk and the concordance lines are manually scrutinized. This process is aided by alphabetically sorting the words that precede or follow the chunk fragment. Through this method, approximate forms resembling native chunks and serving the same organizational purpose are identified. These approximated chunks then serve as a starting point for further searches to uncover additional approximate forms. Among the 5-grams, the chunk "as the matter of fact" (21) and "from my point of view" (19) are the most frequent. Among the 4-grams, the highest frequency is attributed to the chunks "on the other hand" (155) and "at the same time" (133). Lastly, among the trigrams, a preferred choice of ELF users is "so to say" (39), which is an approximation of the conventional chunk "so to speak" (20).

With reference to, among others, the chunk "on the other hand", Carey (2013) claims that there are several approximated forms in ELFA. For instance, "on the other side" is uttered 11 times by seven different users. At the same time, all other approximate forms are spoken once by individual speakers, e.g. "from the other side", "from other side", "on the other hands" or "on the other side of the flip". In WrELFA, approximate forms (3 occurrences) are also found, e.g. "on the one side", or "on one side", which are also noticeable in ELFA. However, no statistical differences between ELF corpora and native corpora are determined. When it comes to the chunk "from my point of view", in turn, Carey (2013) indicates that there are such approximations in ELFA as, for example, "in my view" (5 occurrences) or "in my sense" (3 occurrences uttered by one user). At the same time, "from my point of view" (2 occurrences) and "I my view" (13 occurrences) can also be noticed in WrELFA. Moreover, Carey points out that such approximations seem to be unique to ELF as no approximations are found in the native (MICASE) corpus.

Overall, Carey (2013) discovers that common organization-oriented chunks, such as "at the same time", are equally prevalent in ELF in forms that align with native English standards. This suggests that approximation in ELF exists, but it does not imply that ELF should predominantly be viewed as diverging from native norms or forming its own distinct emerging patterns. Admittedly, approximation primarily concerns organizational chunks that are infrequently encountered, which may indicate that they are weakly entrenched. At the same time, Carey points out that there are no statistically significant differences in the occurrence of approximation in spoken and written data, thereby

suggesting that "the gap between spoken and written ELF may not be that great" (Carey 2013: 226).

2.2.2.2.2. CxG-based studies of written academic ELF data

As indicated in Section 2.2.1.2., CxG is a mature framework and as such able to "bring essential phenomena into the realm of more explicit consideration" (Langacker 2016: 467). Thus, for example, a categorical approach to entrenchment and constructions has been replaced by a gradient perspective through which the complexities of memory and the nuances of constructionhood are (more) adequately explored. Similar tendencies can be observed in CxG-based research into academic ELF[37], where, for instance, entrenchment comes in different strengths, involving fixing as well as approximation, and constructionhood is not determined a priori but established in relation to particular research objectives. These tendencies transpire in the three studies discussed below.

The first study is Vetchinnikova's (2015) analysis of multi-word units which are modeled as Sinclair's (1996, 2004) units of meaning, i.e. form-meaning pairings roughly equivalent to constructions. The aim of the study is to clarify the extent to which ELF use results from "usage-based priming" and the extent to which it relies on "novel combinations of words" (Vetchinnikova 2015: 238). In other words, the question is whether multi-word units are stabilizing patterns emerging from language use or emergent combinations created ad hoc.

To address this research question, Vetchinnikova (2015) analyzes the extent to which five ELF users "recycle experienced language patterns in their own production" (2015: 235). Relying on the production corpus, i.e. the "five students' drafts of Master's theses written over a period of about one year" and the exposure corpus, i.e. "collected academic publications each student referred to in her thesis" (Vetchinnikova 2015: 234), created specifically for the study, the scholar determines the degree of similarity between the multi-word units in the two sets of data using the following procedure. Firstly, both corpora are analyzed by means of the ConGram program (Greaves 2009). To be more specific, Vetchinnikova (2015) configures the internal window span to 4 and sets the minimum frequency to 2, instructing the program to extract all two-word combinations occurring more than once with up to four intervening words. Subsequently, the researcher compares the lists of statistically significant concgrams generated for each student's

[37] For a construction-based study beyond the academic domain see, for instance, Pirc (2013) or Watkowska (2020).

production corpus with their corresponding exposure corpus and establishes the percentage of overlapping concgrams. Concurrently, Vetchinnikova (2015) conducts a comparison by pairing the concgram list from each student's production corpus with the exposure corpus of a different student, where the exposure corpus represents a field of study distinct from that of the student under examination. As the overlap between the production corpus and a non-corresponding exposure corpus is notably lower than the overlap between the production corpus and the corresponding exposure corpus, it seems that the input ELF users are exposed to influences their output, i.e. constructions are likely to emerge from usage.

What is more, in cases where there are discrepancies between production and exposure corpora of the same student, Vetchinnikova (2015) still interprets the results as indicative of emerging preferences, explaining them in relation to the cognitive processes of approximation and fixing (see Section 1.2.2.2.). Regarding approximation, Vetchinnikova notes that one participant in her study uses the patterns such as "amount of deaths" and "number of deaths" in the work, even though the exposure corpus only contains "number of deaths". Vetchinnikova (2015) states that since both "amount" and "number" are used in English as quantifying determiners, the participant simply uses quantifying determiners related to either countable or uncountable nouns in the determiner slot irrespective of the countability of the noun it precedes. Hence, this example illustrates approximation involving the use of rough equivalents of the target norm. In the case of fixing, although a particular pattern, e.g. "the connection between X and Y", does not appear in the exposure corpus of a given participant, Vetchinnikova (2015) points out that the student uses the pattern in a number of different contexts, which indicates an emerging preference.

All in all, Vetchinnikova (2015: 24) sees the modifications introduced into the constructions under study as "oscillations towards fixing or approximation" within units "that have been produced holistically on the idiom principle, even though some of their components may have become more fixed while some may have loosened". Thus, "what might appear to be novel patterning at first sight seems to be a result of natural variation permissible in a unit of meaning rather than a product of creative open-choice assembling" (Vetchinnikova 2015: 25). This observation speaks to the fact that ELF processing is influenced by previous exposure, and is thus usage- and construction-based. In a similar vein but with a more pronounced quantitative focus, Yilmaz and Römer (2020) aim to identify dominant constructions in written academic ELF. Their study relies

on data extracted from the WrELFA corpus and compares ELF data to those taken from the COCA corpus[38] to determine significant differences between ELF and native English constructions. In line with the objectives of the study, constructions are "operationalized as repeatedly used multi-word sequences with an identifiable dominant function, and significantly higher frequency in written academic ELF as opposed to written academic American English" (Yilmaz and Römer 2020: 62).

To identify constructions in the WrELFA corpus, the authors first employ the "key function words to constructions" approach, which entails reliance on, e.g., prepositions, articles, and conjunctions that are statistically more or less frequent in the target corpus (WrELFA) than the reference corpus (COCA), and next establish complex constructions around them.

To be more specific, once statistically significant key function words in WrELFA are revealed and limited to three key prepositions, i.e. "of", "in", and "on", three types of multi-word, preposition-centered units are identified, i.e. contiguous patterns (from 2 to 5 words), collocate lists (with a range of 4 words to the right and to the left of the preposition), and non-contiguous patterns, or phrase-frames, (from 3 to 8 words). In each case, a threshold of approximately 10% is applied, which entails that at least 82 of the 822 texts in WrELFA are considered in the analysis. Subsequently, the results from WrELFA and COCA are compared by means of Rayson's (2016) log-likelihood and effect size calculator to determine significant differences between the corpora (Yilmaz and Römer 2020: 67).

When it comes to the preposition "of", the authors identify 48 patterns of varying complexity which are significantly more common in ELF. The examples are mostly parts of a larger pattern, i.e. "'(DET) N of (DET) N', consisting of an optional determiner, followed by a noun, followed by of, followed by an optional determiner, followed by a noun", the function of which is either to express "the quality or process of a research entity (e.g., the importance of the study) or "the quantity of a research entity (e.g., parts of the thesis, a lot of data, the aim of this study)" (Yilmaz and Römer 2020: 71–72). Importantly, the two meanings of the polysemous construction are essentially based on open-class words. As Yilmaz and Römer explain, since "[f]requent nouns in the first N slot include analysis, number, and results, while study, work, and data are among the most frequent nouns in the second N slot", three semantic categories can be identified:

[38] COCA stands for the Corpus of Contemporary American English, representing data pertaining to native users.

"'research clusters', e.g., the results of the study", "'quality clusters', e.g., the presence of, use of the", and "'quantity clusters', e.g., part of the, a number of" (2020: 70).

With reference to the preposition "in", Yilmaz and Römer (2020: 73) state that the form "'(DET) N V-ed in (DET) N', consisting of a determiner, followed by a noun, followed by the past participle of a verb, followed by in, followed by a determiner, followed by a noun (e.g., the research presented in this paper, all data represented in this thesis)", expressing "how a certain means is utilized in a research product or entity" is significantly more frequent in WrELFA than COCA. In relation to the preposition "on", Yilmaz and Römer (2020: 75–76) identify four characteristic ELF constructions, i.e. "V on (the) N", e.g. based on findings, "N on (the) N", e.g. impact on the development, "(DET) N of N on (DET) N", e.g. the effects of preservation on methane, and "(DET) N is V-ed on (DET) (N)", e.g. this conclusion is based on the author's results. Nevertheless, the discussion of their function is available only in relation to the last two examples. More specifically, while "(DET) N of N on (DET) N" refers to the impact of a research endeavor on an entity, "(DET) N is V-ed on (DET) (N)" denotes a research procedure or entity and its connection to a source.

The authors conclude that the constructions identified in the study are not unconventional or unidiomatic as they are also frequently used by academic native English writers. However, according to the analysis of WrELFA and COCA, these constructions are notably more common in academic ELF than in academic native English. Moreover, it is noted that certain ELF constructions "are not fully schematic in that writers use them with a limited set of lexical items. This finding is similar to observations made in studies on construction development in first language acquisition, such as Tomasello (2003), which show that learners initially resort to 'constructional islands' before expanding their repertoire to more schematic and productive constructions" (Yilmaz and Römer 2020: 78). Finally, the authors suggest that their study could be refined and extended if constructions were explored with reference to, for instance, academic disciplines, text types and writers' L1s. Additionally, "[a] second modification of the approach could involve lowering the frequency and range thresholds we worked with which would allow us to capture potentially less conventional constructions which are overall less frequent and less well dispersed, and which are potentially approximations of conventional constructions (Mauranen 2012)" (Yilmaz and Römer 2020: 79–80).

Yilmaz (2020) attempts to include at least some of the above suggestions in his analysis of written ELF constructions, i.e. "multi-word sequences with discourse-functional properties in three corpora of academic writing from 50 disciplines in the social and natural sciences". To be more specific, Yilmaz aims to determine whether variables such as edited and unedited academic text types, disciplines and native languages (similects) of the authors influence the results concerning the dominant written ELF constructions. The work relies on data derived from the sub-corpus of WrELFA, i.e. the corpus of unedited research papers. Additionally, two corpora, one including published research articles by non-native and the other by native users of English, are compiled by Yilmaz to facilitate comparisons between, e.g. constructions distinctive of unedited ELF, edited ELF, and edited ENL. Following a procedure similar to that presented by Yilmaz and Römer (2020), the three corpora are systematically analyzed.

First, 26 prepositions are identified as significant across the corpora, with 12, i.e. "of", "for", "to", "across", "after", "over", "within", "through", "in", "at", "by", and "on", functioning as key. Then, corpus-based investigation of multiword units leads to the discovery of 19 constructions utilizing the 12 key prepositions. Specifically, the prepositions "of" and "in" enable the identification of six and three constructions, respectively, while the remaining prepositions develop into one construction. The following six form-meaning pairings represent the constructions identified across corpora: "N of (NP)" – abstract or concrete part-whole relation between two entities, "N of NP in (NP)" – abstract or concrete part-whole relation between two entities at a location, "N for (NP)" – generally abstract and sometimes concrete intentional relation between two entities, "N to (NP)" – mostly abstract and rarely concrete directional relation between two entities, "within N" – spatial containment and abstract contextualization, "(NP) V-be V-ed by (NP)" – prepositional passive used with verbs that denote: discourse construction, research processes, or research findings (Yilmaz 2020: 92).

At the same time, it is important to note that Yilmaz (2020: 63) defines the meanings of prepositions largely in consonance with their emobodied understandings, such as "spatial and metaphorical senses of containment or enclosure of the preposition 'in' (Dirven, 1993; Radden & Dirven, 2007)" or the senses of the preposition "by" which, "aside from a temporal and spatial connectedness, (…) can be used to also denote a means, cause, or specification of abstract and quantitative properties (Dirven, 1993; Radden & Dirven, 2007; Sinclair, 1996; Strauss et al., 2018)" (Yilmaz 2020: 84). As a

result, while the meanings of the constructions identified by Yilmaz (2020) are still considerably scaffolded by discourse-functional meanings of their open-class components, which is particularly clear in the case of the last pattern quoted above, i.e. "(NP) V-be V-ed by (NP)" paired with detailed semantics of the verb, but also visible in the complex hierarchies underlying the abstract and concrete domains proposed for the noun,[39] embodied meanings of closed-class elements are also prominent.

When it comes to statistical comparisons, the two ELF corpora exhibit significantly higher frequencies compared to the native English corpus for certain constructions, e.g. "after N". Conversely, the ENL corpus demonstrates notably higher frequency counts than both ELF corpora for, among others, "ADJ of (NP)". Regarding the disparities between the two ELF corpora, only "over N" and "DET of (NP)" are found to be significantly more frequent in ELF (edited) than in ELF (unedited). While these tendencies indicate important differences between the three corpora in terms of usage of particular constructions, a more general propensity can also be noted. As Yilmaz (2020: 165) puts it, "[t]he key constructions in ELF generally denoted abstract meanings and processes commonly found in academic discourse. On the other hand, those in ENL made use of concrete meanings for specifying quantities and properties of entities".

With reference to similects and disciplines within and across corpora, Yilmaz (2020) claims that the two variables influence the way in which a given construction is applied. For instance, the analysis shows that Finnish and Czech authors use the "N for (NP)" pattern in a manner which is not shared by ELF users from the other L1 backgrounds. Also, there is a difference in the use of "N to (NP)" between the disciplinary domains of social sciences and humanities and natural sciences. Hence, Yilmaz (2020) states that research concerning ELF constructions should take into account that their characteristics cannot be simply regarded as a set of uniform features of written academic ELF shared by writers from various linguistic and disciplinary backgrounds.

At the same time, though, it is also possible to determine universal ELF constructions, i.e. patterns that reflect

> general characteristics of written academic ELF that transcend the observed similectal and disciplinary differences. The identified ELF 'universals' included: the common use of conventional academic constructions such as nominals and passives with high-frequency

[39] To be more specific, the abstract domain is further divided into five subdomains encompassing general, process, framing, cognitive, and research nouns whereas the concrete domain includes quantity nouns and nouns referring to humans, animals, and entities. Due to the very low frequencies within subdomains, only the main domains, i.e. abstract and concrete, are employed in the study.

> prepositions, as well as abstract meanings for general and discursive purposes; limited use of constructions around low-frequency prepositions with remarkable variation especially in SCI [natural sciences] writing; and distinctive expressions of evaluation and voice (Yilmaz 2020: 166).

All in all, Yilmaz's (2020) study uncovers important constructional characteristics of written academic ELF, such as its tendency towards simplification and complexity, conventionality and variation, universality and relativity, schematicity and specificity, and reveals important insights into an interplay between L1s, disciplines and constructions. Still, given the fact that written academic ELF is "an underresearched type of language" (Yilmaz and Römer 2020: 80), further studies into the nature of the relations between linguistic constructions and their contexts are encouraged. Such explorations, as indicated by Yilmaz (2020: 170–171), should not only refine and expand the impact of similects and disciplines upon constructions but also delve into other aspects of context, such as those envisaged by a textographic approach (Swales 1998), so that written academic ELF constructions can be adequately described as emerging not only from one category of resources, e.g. linguistic, or one group within a community, e.g. authors of research articles, but from a richer mosaic, or tapestry (see Section 2.1.1.), including, for instance, paralinguistic resources and editors of academic journals.

2.3. Concluding remarks

The overall aim of Chapter 2 was to present academic ELF from a usage-based perspective. Thus, in Section 2.1., an interplay between the academic domain and ELF was discussed. Defined as a network of overlapping groups whose fuzzy contours could be delimited by, for instance, geography and discipline, and whose activities, such as writing, editing and publishing research articles, could be successfully conducted through English, academia was seen as decidedly shaped by the international factor. Against this backdrop, the role of some influential ELF users, e.g. non-native English authors of research articles, in developing the academic domain was highlighted, while the impact of other members of the academic ELF community, e.g. non-native English editors of (peripheral) journals, was seen as less-readily acknowledged. Nevertheless, expertise-based academic ELF could be felicitously defined as a repertoire of linguistic resources co-constructed and shared by a heterogeneous and multi-layered community and thus appropriately researched through a usage-based approach.

Section 2.2., in turn, showed that the apparent suitability of a usage-based approach for researching academic ELF was not fully recognized and applied. Two grammars, i.e. LUG and CxG, were seen as particularly valid but clearly underexplored models through which emergent and emerging form-meaning pairings could be captured. To adequately reveal the potential of LUG and CxG for researching academic ELF, their theoretical foundations were first laid and then extended by means of empirical details. To be more specific, the first stage depicted LUG and CxG as complementary approaches which, when integrated, could appropriately explain both online and offline nature of linguistic experience. In other words, it was argued that the (mostly) emergent patterns revealed through LUG and the (predominantly) emerging, or even emerged, constructions uncovered through CxG were, in fact, different phases of the same phenomenon and could thus be placed along the cline of emergence and explained with reference to a number of related continua, e.g. working-/long-term memory, perceptual/usage-based chunking, intuition (innateness)/learning (experience), linearity/hierarchy. The second stage, in turn, discussed the two main databases of academic ELF, i.e. ELFA and WrELFA, and provided detailed descriptions of four LUG-based and three CxG-based studies of spoken and written academic ELF. With reference to data collection, it was observed that clear criteria, i.e. compilation principles, should be defined if a usage-based approach to academic ELF was to rely on stable sets of data, particularly in the form of new corpora constructed to satisfy the growing need for capturing more naturally-occurring academic ELF data. To be more specific, newly-constructed ELF databases, especially pertaining to written data, should facilitate further research into more and less established genres, taking into account their different layers, e.g. linguistic and paralinguistic. With reference to research design, it was noted that within LUG-oriented analyses further studies should be devoted to exploring the relationship between online and offline chunking through, for instance, tracing the recurrence of emergent clusters, particularly within written data. In other words, more attention should be paid to the segmentation of incoming information into spatial groups and the frequency of such groups so that emerging ELF-specific constructions could be revealed. Within CxG-oriented studies, in turn, it was observed that future analyses should attempt to build stronger bridges between cognitive processes at individual and collective levels, low- and high-frequency patterns, assumed (emerged)

and identified (emerging) constructions[40], embodied and discursive dimensions of meaning and linguistic and paralinguistic components of form. In other words, the potential provided by the gradient view upon constructionhood should be fully exploited so that the complexity of written academic ELF could be better understood, especially in relation to constructions going beyond the (currently analyzed) sentence level.

In fact, a robust constructionist approach seems vital not only for uncovering the nature of written academic ELF but of ELF in general since a number of observations made in relation to the former were already articulated in Chapter 1. Among others, ELF, both general and academic, would benefit from a fully-fledged embodied perspective through which the apparently neglected cognitive processes of Gestalt-base grouping and figure-ground segregation (see Section 1.2.2.), which underlie human thought and are thus readily available to all ELF users (see Section 1.2.2.2.), and which underpin the emergent-emerging continuum (see Sections 2.2.1.1. and 2.2.1.2.), could be systematically incorporated into a construction-based ELF research. Such a perspective is undoubtedly offered by Cognitive (Construction) Grammar, discussed in detail in Chapter 3.

[40] Potentially, this implies the importance of a clear differentiation between the use of explicit norms, e.g. dictionaries, as a starting point of a given analysis and the subsequent determination of new regularities (see Section 3.1.2.2.1.)

Chapter 3
Cognitive (Construction) Grammar as a usage-based approach

3.0. Outline

The aim of Chapter 3 is to present Cognitive (Construction) Grammar (henceforth also C(C)G) as a usage-based model through which the nature of constructions in academic ELF can be adequately captured. The explanatory potential of this usage-based approach is presented through its cognitive and constructionist commitments, introduced in Section 3.1. and elaborated in Sections 3.1.1. and 3.1.2., respectively.

With reference to the cognitive commitment, emphasis is laid upon Gestalt-based construal operations which constitute a set of cognitive processes uniquely highlighted in C(C)G. The importance of Gestalt principles for C(C)G is established in three steps. First, Section 3.1.1.1. outlines an interplay between cognition and Gestalt organization, next Section 3.1.1.2. concretizes Gestalt principles as cognitive, or construal, operations, a.k.a. basic concepts, and finally, Section 3.1.1.3. establishes basic concepts as foundational to the conceptual system and thus to the architecture of linguistic constructions.

This architecture is discussed in detail in Section 3.1.2., where the constructionist commitment of C(C)G is presented. To be more specific, Section 3.1.2.1. defines constructions in accordance with C(C)G, while Section 3.1.2.2. characterizes form-meaning pairings along the continua of fixedness-novelty (Section 3.1.2.2.1.), specificity-schematicity (Section 3.1.2.2.2.), and simplicity-complexity (Section 3.1.2.2.3.).

The chapter ends with a summary in which the key assumptions of Cognitive (Construction) Grammar are highlighted and possible routs for its further application are outlined.

3.1. Cognitive (Construction) Grammar as a usage-based model

As stated in Section 1.2.2., a usage-based approach views linguistic knowledge as (more or less) established form-meaning pairings, i.e. linguistic units or constructions, emerging from an interplay between language use and cognitive processes. In other words, a usage-based perspective gives "substantial importance (…) to the actual use of the linguistic system and a speaker's knowledge of this use" (Langacker 1987: 494), while "[t]he assumptions made about mental abilities and cognitive processing are both minimal and

relatively non-controversial" (Langacker 1999: 93). On a more detailed note, a usage-based approach, such as Cognitive (Construction) Grammar[41] (Langacker 2009)[42], can be discussed with reference to its cognitive (Section 3.1.1.) and constructionist (Section 3.1.2.) commitments.

3.1.1. The cognitive commitment of Cognitive (Construction) Grammar

Situated within cognitive linguistics[43], or the cognitive linguistics enterprise, i.e. "an approach that has adopted a common set of core commitments and guiding principles, which have led to a diverse range of complementary, overlapping (and sometimes competing) theories" (Evans et al. 2007: 3), Cognitive (Construction) Grammar embraces its two foundational claims: the generalization commitment and the cognitive commitment. While the former is "a dedication to characterizing general principles that apply to all aspects of human language" (Evans 2012: 3), the latter "represents the view that principles of linguistic structure should reflect what is known about human cognition from other disciplines, particularly the other cognitive sciences (philosophy, psychology, artificial intelligence and neuroscience)" (Evans and Green 2006: 40–41).

However, C(C)G is not only an integral element of cognitive linguistics but also "the most detailed theory of grammar to have been developed within cognitive linguistics and to date has been the most influential" (Evans and Green 2006: 480). The unwavering impact of C(C)G is caused by, among others, its long-lasting presence on the cognitive scene (Langacker 1987, 1990, 1991, 2001, 2008, 2017a, 2017b) and the unchanged nature of its fundamental claims. These fundamentals not only depart from competence-based models (see Section 1.2.1.) but also set C(C)G from at least some of the other construction-based approaches (for an overview see Hoffmann and Trousdale 2013) in at

[41] Cognitive (Construction) Grammar is a term introduced by Langacker (2009) in an attempt to show points of convergence between his model of Cognitive Grammar (CG) and Goldberg's (1995, 2006) conception of Construction Grammar (CxG) (see Section 2.2.1.2). Thus, while the term is essentially used here as a rough equivalent to Cognitive Grammar, the parenthesized reference to constructions is maintained to indicate potential expansions of CG.

[42] Relatedly, Ungerer (2023: 5) notes Langacker's "immense impact on constructionist theorizing, even though he is perhaps not a Construction Grammarian in the narrow sense. Nevertheless, many of his ideas have been integrated into Construction Grammar, and his own framework of Cognitive Grammar is sometimes regarded as a subtype of the former (e.g., Langacker 2005)".

[43] Unless otherwise indicated, the adjective "cognitive" is used here to refer to "second-generation" (Lakoff and Johnson 1999: 77–78) theories in the cognitive sciences, i.e. such that do not see the mind as based (solely or predominantly) on abstract, propositional representations and as largely independent from specific brains, bodies, and sensory modalities in processing information. Hence, cognitive linguistics is also understood as second-generation cognitive linguistics, i.e. an approach complementary to the Chomskyan model (see Section 1.2.1.).

least two ways. The first one is the claim that grammar is symbolic, and thus inherently meaningful, and the other is the argument that "[n]ot only is it meaningful, it also reflects our basic experience of moving, perceiving and acting on the world. At the core of grammatical meanings are mental operations inherent in these elemental components of moment-to moment living" (Langacker 2008: 4–5). These mental operations can be divided into memory-based, interaction-based and perception-based (see Section 1.2.2.).

Memory-based, or learning, processes involve: automatization, i.e. mastering a structure which, through repetition, "undergoes progressive entrenchment and eventually becomes established as a unit" (Langacker 2008: 16), schematization, i.e. "the process of extracting the commonality inherent in multiple experiences to arrive at a conception representing a higher level of abstraction" (Langacker 2008: 17), and categorization, i.e. "a special case of comparison, obtaining when the standard [of comparison] represents an established unit and the target (at least originally) is novel" (Langacker 1999: 94).

Interaction-based processes are those that lead to knowledge alignment. As Langacker explains, "[e]ach of us has our own mental universe, obviously, but since it develops largely through social interaction in a culture, it's usually similar enough from person to person to serve as a kind of common ground, at least portions of it can. So this overlap provides what I call a shared conceptual substrate" (2017a: 310). In other words,

> the units that different speakers learn are similar enough to allow effective communication. No one is going to have exactly the same language as anyone else. In fact, the differences may be very substantial, even within a small speech community. But there is enough similarity, because it all arises through and continues through interaction, that we can communicate (Langacker 2017a: 112)

In fact, the importance of social interaction is reflected in C(C)G in a number of ways. Firstly, "a unit is part of a language only to the extent of being conventional in a certain speech community. Next, units emerge from language use, which encompasses "the very fact that the speaker and hearer are interacting by using the language in question. Hence the ground (the interlocutors, their interaction, and its circumstances) figures at least peripherally in the import of every unit" (Langacker 2016a: 469).

Perception-based, or conceptualization, processes are those "inspired by general psychological research on vision" and strongly influenced by principles of Gestalt psychology, such as figure-ground segregation (Diessel 2017: 10). Apparently, unlike memory- and interaction-based processes which are present in most usage-based approaches, the link between (Gestalt) perception and cognition is particularly emphasized in C(C)G. As Dirven (2005: 17) puts it, C(C)G is thus firmly situated within

"[a] gestalt-psychology-based strand, initiated and explored by Talmy, and worked out in greater detail by Langacker". A similarly distinctive connection between the Gestalt approach and C(C)G is noticed by Gobet (2017: 251), who sees Langacker's model as a comprehensive transfer of Gestalt principles of, among others, grouping and figure-ground segregation to grammar.

Due to its unique position in C(C)G, Gestalt-based processes are discussed in detail below. To be more specific, Section 3.1.1.1. situates Gestalt organization within the cognitive architecture adopted in C(C)G and highlights its perceptual underpinnings, Section 3.1.1.2., in turn, focuses on the conceptual side of Gestalt principles and presents them as basic cognitive abilities, alternatively theorized as basic concepts. Finally, Section 3.1.1.3. shows how the conceptual system develops from basic, including Gestalt, concepts into an intricate knowledge network.

3.1.1.1. Cognition and Gestalt organization

As already noted above, the cognitive commitment of the cognitive linguistics enterprise, including C(C)G, means that language should be described in accordance with what is known about cognition from, e.g., psychology and neuroscience. Broadly speaking, this knowledge comes as two viewpoints, i.e. the classic cognitive strand, which maximizes the role of computational processes and sees the brain as the sole seat of cognition, and the post-cognitive strand, which is committed to "minimiz[ing] (...) the need for rule-guided inferences (...) and, hence, the need for computation" (Shapiro and Spaulding 2021: online), and maximizing the need for incorporating the body and its surroundings in interpreting cognition (Barsalou 2020)[44].

Cognitive linguistics follows the second strand and sees "language as a reflection of embodied cognition, which serves to constrain what it is possible to experience, and thus what it is possible to express in language". In other words,

> [c]ognitive linguists argue against the view that language is pre-specified in the sense that grammatical organisation is mapped out by an innate 'blueprint' for grammar, and semantic organisation by a set of semantic primitives. Instead linguistic organisation is held to reflect embodied cognition (...), which is common to all human beings (Evans and Green 2006: 63–64).

[44] Simultaneously, Barsalou (1999) emphasizes that definitions of cognition vary, with researchers classifying not necessarily identical processes as belonging to cognition. Still, memory, language and thought are rather consistently viewed as forming cognition, or at least its higher level.

One of the ways in which embodied cognition constrains experience and the conceptual system based on this experience is through the principles of Gestalt perception (Evans and Green 2006: 65–68). The mechanics of this interplay are discussed below in two steps; first from the angle of embodied cognition (Section 3.1.1.1.1.) and then from the perspective of Gestalt principles (Section 3.1.1.1.2.).

3.1.1.1.1. Embodied cognition

While the idea of embodied cognition is generally embraced in the framework of cognitive sciences, the interpretations of the notion are far from consistent. Wilson (2002), for instance, distinguishes six views upon embodied cognition, Gallagher (2011) puts forward five senses of embodied cognition, and Shapiro and Spaulding (2024) distinguish three themes around which embodied cognition can be discussed, signaling, at the same time, its connection to the other "Es" in cognitive science research, i.e. embedded, extended and enactive. Overmann and Malafouris's (2018) classification, in turn, can be used to neatly illustrate the complex relationship between the various interpretations of cognition. The list encompasses the following E-domains "engaged during cognition" (Barsalou 2020: 2): embodied (recognizing the role of human physiological and sensorimotor capacities in negotiating an individual's functioning in the world as well as forming mental representations of this experience), embedded (highlighting the impact of the natural and sociomaterial environment upon behavioral and psychological responses of members of communities, including their conceptual systems and languages), enactive (accentuating the interactive nature of cognition and equating thinking with doing), and extended (seeing the mind as externalizing its functions through recruiting and incorporating resources and processes outside the brain, such as artifacts or technologies)[45].

Relatedly, embodiment is variously interpreted in cognitive linguistics. Lakoff (1987: 266–267), for instance, defines embodiment in a broad manner, i.e. as "our collective biological capacities and our physical and social experiences as beings functioning in our environment". Lakoff and Johnson (1999), in turn, show three levels of embodiment, i.e. the neural, the phenomenological and the cognitive unconscious. For

[45] Moreover, Overmann and Malafouris's (2018) classification encompasses dynamical cognition (viewing the mind as a complex system made up of a number of related variables interacting on many levels), and distributed cognition (involving networks of interacting agents and objects engaged in situated practices and problem–solving activities).

Gibbs (2017: 450), "[e]mbodiment refers to the ways persons' bodies and bodily interactions with the world shape their minds, actions, and personal, cultural identities". Thus, "[t]he study of embodiment demands recognition that thought and language arise from the continuous dynamic interactions between brains, bodies, and the world". A similar interpretation is proposed by Langacker (2014: 30), who observes that "[t]he brain is an integral part of the body, which in turn exists in a world with which it interacts at many levels. Brain, body, and world all have specific structural properties that shape and constrain their interaction and thus the nature of human experience. This is the notion of embodiment".

At the same time, Langacker (2014: 32–34) explains the relation between embodied and embedded cognition "as a dialectic engagement manifested at successively higher levels of conceptual organization". This link can be understood "metaphorically in terms of embodied cognition providing raw materials that are shaped by culture into specific forms, [i.e.] embodiment creates the potential for culture, which constitutes a particular realization of that potential". Hence, "cognition is universally embodied and culturally embedded" (Yamaguchi et al. 2014: 8). Likewise, Kövecses (2015: 75) observes that "[t]he meaning–making organs of the body and brain are shared universally and thus they do not belong to particular cultures. They are thus responsible for universal meanings — meanings shared by all groups of people (…). However, (…) both the body and the brain may be imbued with culture–specific meanings in particular cultures"[46].

The universality entailed by embodiment is reflected in a set of basic cognitive abilities which, according to C(C)G, adequately characterize the notion. As Langacker (2014: 30) puts it, while embodiment can be theorized in many ways, in practice "the notion is mostly limited to the physical aspects of experience" and ultimately captured by a number of, presumably innate (Langacker 1999: 171), processes which can be "readily demonstrated at the level of basic perception" (Langacker 2008: 104).

3.1.1.1.2. Gestalt principles

An interplay between cognition, e.g. memory, knowledge and language (Barsalou 2020), and perception, e.g. vision, audition and gustation[47], is fundamental to embodiment and

[46] Relatedly, Kövecses (2015: 26) proposes the notion of cognitive styles, i.e. "ways in which members of a group employ the (otherwise universal) cognitive processes available to them".
[47] According to Barsalou (1999), perception encompasses vision, audition, haptics, gustation, olfaction, proprioception, kinesthesia and introspection.

hence to C(C)G. In other words, perceptual structures, "[r]ather than being merely input and output devices", are central to cognition (Pecher and Zwaan 2005: 1–2) and thus play a role in, among others, storing and retrieving information from memory (see, e.g., Solomon and Barsalou 2001) and understanding language (see, e.g., Gallese and Lakoff 2005; Borghi and Binkofski 2014).

Against this background, where "perception and cognition share common neural systems, they function simultaneously in the same mechanisms and cannot be divorced" (Barsalou 1999: 603)[48], it is only natural that Gestalt principles, which "allow unconscious perceptual mechanisms to construct wholes or 'gestalts' out of incomplete perceptual input" (Evans and Green 2006: 65), should play a role in, among others, language-related processes. Still, to convincingly reveal the embodied nature of these processes, their perceptual underpinnings need to be duly highlighted. Thus, in consonance with Langacker's (2008) proposal to distinguish between grouping, i.e. region formation, and reification, i.e. region interpretation, Gestalt organization is discussed as a two-stage process below, focusing on examples from the domain of visual perception.

3.1.1.1.2.1. Grouping

As stated by Johansson and Ulrich (2024: online), "[t]he human visual system is highly fluent in parsing optical input into structured and meaningful units, commonly termed perceptual groups". This grouping fluency is based on "the fact that observers perceive some elements of the visual field as going together more strongly than others" (Wagemans et al. 2012: 1193), which is due to the presence of conducive conditions (see Figure 3), such as, among others, proximity, similarity, common region, continuity or closure (Rock and Palmer 1990).

[48] Importantly, the claim that cognition and perception are intertwined does not mean that cognition should be reduced to perceptual (or sensory-motor) representations. In fact, strongly embodied theories seem dominated by approaches which "leave room for higher-level modal, crossmodal, or even heteromodal representations and often acknowledge that a degree of abstraction takes place within and between modalities (Simmons & Barsalou, 2003; Vigliocco et al., 2004). In general, researchers are moving in the direction of weak embodiment (Barsalou, 2016; Pulvermüller, 2013)" (Dove 2020 et al.). Günther (2016: 90) describes approaches based on weak embodiment as such that "do not fully equate perception and cognition but understand them to be stages or aspects of the same process". In particular, such approaches maintain "that knowledge, and thus also meaning, is grounded in concrete perceptual experiences, but that the information gained from these experiences becomes interpreted (Barsalou 2012: 239–240) and integrated into an individual's existing knowledge base. In this course, the memory traces of specific experiences undergo a range of processes of analysis, generalization and abstraction".

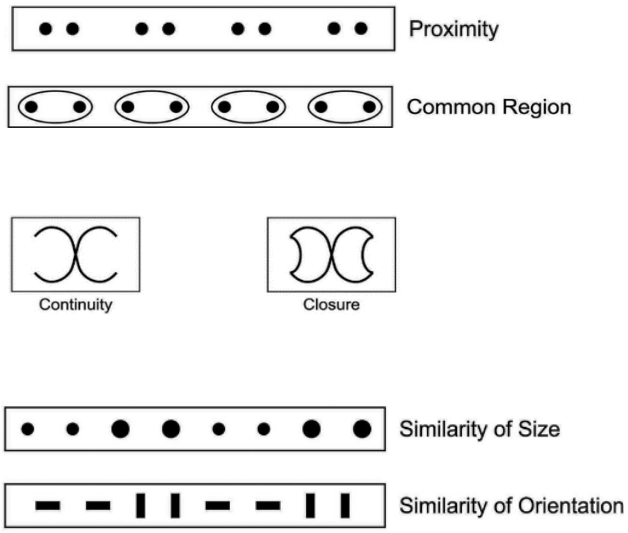


Figure 3. Gestalt principles (adapted from Wagemans et al. 2012: 1980)

These conditions, or Gestalt factors, are features to which humans naturally respond and on which the brain spontaneously imposes an organization (van Leeuwen et al. 2011: 38). Relatedly, there is "evidence about 'features' separated in the visual cortex in which each represents a different perceptual dimension such as color, orientation and movement. The organization of the elements [i.e. grouping] is presumably carried out based on factors such as similarity, proximity, contiguity, direction and similarity" (Giesteira et al. 2010: 2). In other words, at least some grouping principles are ultimately rooted in neurophysiological structure (Spillmann 2009), and thus universally available[49]. For instance, Quinn et al. (2008: 296) argue that "a considerable degree of perceptual organization is present in the initial months of life and does not await an extended period of learning the statistics of visual images". Moreover, Wagemans et al. (2012: 1193) discuss research results which suggest "that infants are capable of grouping visual elements into unitary structures in accord with a variety of both classic [e.g. proximity and similarity] and modern [e.g. common region] organizational principles". On the other hand, however, Wagemans et al. (2012: 1207) show that "not all grouping cues are readily

[49] At the same time, it should be noted that grouping principles at least partly overlap with another universally available set of determining factors, i.e. principles of figure-ground interaction. Thus, as Wagemans et al. (2012: 1208) conclude, grouping and figure–ground interaction are related perceptual processes which can be arranged "on a graded continuum" and described with reference to such shared parameters as, e.g. proximity and closure.

available to young infants and that there is a protracted developmental trajectory for some perceptual organization abilities, even those that appear to emerge during infancy". A compromising view is that certain grouping principles, i.e. position–based Gestalt principles, e.g. proximity, common region and continuity, are likely to have "neuronal correlates" (Spillmann 2009: 1511), while similarity–based principles are less likely to be intrinsic to the brain (Quinn et al. 2008). In other words, while both proximity and similarity are the "primary" factors which encourage grouping (Langacker 2008: 104), there seems to exist "the proximity advantage" (Johansson and Ulrich 2024) by which proximity-based grouping (encouraged by closeness in space) occurs faster and is more accurate than grouping based on similarity (encouraged by shared featural properties such as color)[50].

Another encouragement to grouping is provided by the universal tendency of the human perceptual system through which a more (rather than less) integrated entity is perceived, i.e. "an entity that seems 'to define a basic form or skeleton, a more perfect original image, from which the real object only seems to be diverted through certain distortions, deletions or additions'" (Stadler 2020: 220). This inclination is known as a tendency towards Prägnanz[51], which Barsalou (1999: 586) illustrates in the following way: "[w]hen a linear series of points is presented visually, an underlying line is perceived. As a result, the stored perceptual information goes beyond what is objectively present, representing a line, not just the points"[52] (see Figure 4).

[50] Still, as Langacker (2008: 105) observes, "grouping (…) occurs at multiple levels" and thus alternative groupings or groupings within groupings are to be expected. Likewise, Todorovic (2008: 5345) notes that "[i]n case two (or more) principles apply for the same input, and they favor the same grouping, it will tend to become strengthened; however, if they disagree, usually one wins or the organization of the percept is unclear". While the conundrum "has been addressed to some extent in the literature (e.g. see Kubovy & van der Berg, 2008), the significant theoretical problem of how to predict which principle will win in which circumstances remains to be worked out in much more detail".

[51] Apparently, this preference exists since "evolution may have built simplicity mechanisms into the visual system as a heuristic for likelihood" (Palmer 2003: 9). In other words, "[s]traight lines, continuous contours, symmetrical shapes are ubiquitous properties of natural objects. However, objects are rarely presented in their entirety. To make up for any stimulus occlusions or distortions, neuronal mechanisms may have evolved that strive to perceptually rectify crooked lines, fill in gaps, and complete patchy surfaces (...), thereby restoring the stimulus to its original state" (Ehrenstein et al. 2003: 434).

[52] Langacker (2008: 105) calls this grouping principle "the recognition of familiar configurations".

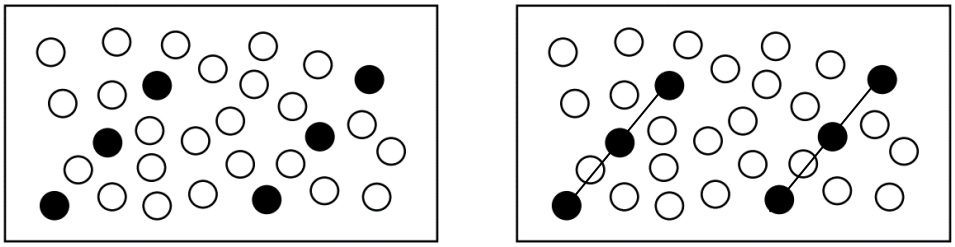


Figure 4. Prägnanz (adapted from Langacker 2008: 105)

In fact, simplification is fundamental to Prägnanz, also known as the simplicity principle, which "holds that the visual system chooses the simplest interpretation, that is, one that can be specified by a minimum number of descriptive parameters" (Wagemans et al. 2012: 1236), thus arriving at a pattern which is "as simple, orderly, balanced, unified, coherent, regular, etc as possible, given the input" (Todorovic 2008: 5345). In other words, Prägnanz is characterized by "intuitive clarity, singularity and internal order that can (…) lead to pure homogeneity" as opposed to a Gestalt which lacks "any uniformity such that it blurs with or disappears entirely in its environment" (Sadler 2020: 221). Moreover, Prägnanz is integral as opposed to a Gestalt in which "something can be absent, missing, there might be a hole, there might be too little of something there. There can also be too much, a superfluity, a growth, an alien body" (Smith 1988: 63–64). Next, Prägnanz is "independent, disconnected, (…), disengaged from surroundings" (Edwards 2012: 13). Finally, simplicity and complicatedness, or complexity[53], can be taken as a "gradual Prägnanz aspect".

In fact, a gradual scale between the extremes of maximal and minimal goodness can be applied to all Gestalt properties which, in the case of Prägnanz, adopt maximal values, as illustrated by the images in the right-hand column of Figure 5 below.

[53] In fact, Stadler (2020), following Rausch (1966), distinguishes between complicatedness and complexity, seeing the former as a negative Gestalt property and the latter as a positive one. Wagemans et al. (2012), on the other hand, uses simplicity as a prägnant characteristic and complexity as its opposite. In view of this conundrum between complicatedness and complexity, the solution adopted here is to use the term non-simple to describe an unfavorable Gestalt property which makes an object "intuitively and intellectually hard to grasp" (Stadler 2020: 221).

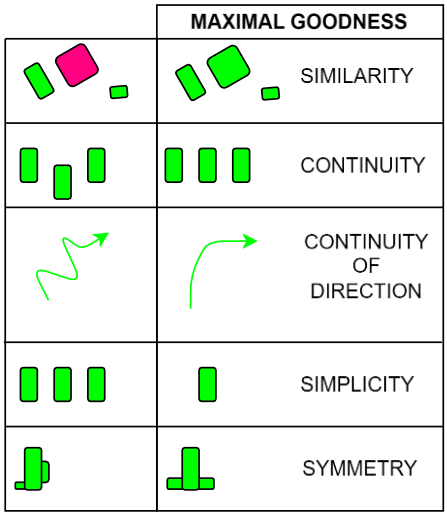


Figure 5. Maximal goodness (adapted from Duke 1999: 20)

While the universality of grouping principles, and the properties they entail, is an important aspect of Gestalt organization, particularly in the context of embodiment ultimately defined through a set of innate cognitive abilities (see Section 3.1.1.1.1.), grouping is also clearly facilitated by experience and thus embedded. For instance, as Todorovic (2011: 6), following Wertheimer (1923), observes,

> the existence of a Gestalt grouping law based on past experience or habit (…) [can be] manifested in reading (…). Thus, we are likely to perceptually subdivide a sequence of characters such as '314cm' into two parts, '314' and 'cm', rather than, say, into '31' and '4cm', presumably because we have learned to differentiate letters and numbers, and how they are usually combined.

Likewise, "[t]he spontaneity and ease of the standard, dominantly perceived organization of the strokes into letters, is plausibly mainly due to past experience, that is, to our familiarity with words as written in the script form of the Roman alphabet. This particular organization might not occur for observers lacking such familiarity" (Todorovic 2008: 5345). All in all, both universal and culture-specific grouping principles lead to the emergence of groups, also referred to as unitary forms, (sub)wholes, or Gestalts (Todorovic 2008), and facilitate their description with reference to a number of properties.

3.1.1.1.2.2. Reification

According to Langacker (2008: 105), reification is the "capacity to manipulate a group as a unitary entity for higher-order cognitive purposes". To illustrate, each line in Figure 4 above, "is reified by virtue of being treated as a single element in the perception of parallelism [i.e. symmetry], and in any other conception invoking them as such (e.g. in

counting them [i.e. assessing simplicity], comparing their length [i.e. assessing similarity], or observing their slope [i.e. assessing continuity])".

In fact, reification entails that groups, or dynamic whole–part configurations (Stadler 2020)[54], can be described not only with reference to symmetry, simplicity, similarity and continuity but also to a number of other Gestalt properties, e.g. homogeneity, independence, and closure. These properties, however, are far from clear-cut. Similarity, continuity, and closure, for instance, are all internally complex categories, with many types of similarity (see Figure 3), at least one elaboration of continuity (Figure 5), and closure instantiated through either an absence of gaps or a presence of "a satisfactory conclusion" (Meyer 1961: 130), i.e. through a tendency of the perceptual system "to close gaps and complete unfinished forms" (Chang et al. 2002: online). Moreover, Gestalt properties tend to (partly) overlap, e.g. continuity is taken as a kind of closure (Meyer 1961). On top of it, as Todorovic (2008: 5345) observes, operationalizing Gestalt properties may sometimes be problematic, i.e. "whereas it may be relatively simple to point out the presence of continuity, closure, etc, it is more difficult to establish what exactly makes a pattern visually good, simple, unified etc". Thus, it seems advisable to limit Gestalt properties in a motivated way.

One such way is suggested by Radden and Kövecses, who claim that "[a]n essential requirement of any gestalt is that it has clearly delineated boundaries" (2007: 22)[55], which entails that a good, or prägnant, Gestalt should, or even must, be shaped or bordered (Stadler 2020), rather than shapeless or borderless or, as stated above, blurring with or disappearing in its environment. Interestingly, this requirement is in line with Pinna's (2012: 279) observation that

> shape defined by boundary contours is more important for visual organisms in terms of adaptive fitness than are [for instance] color, shading, and lighting. In fact, it warns the presence of an object. The primacy of the boundary contours is also phenomenally related to the fact that contours alone can signal information also about surface color and depth as proposed by Rubin.

[54] Gestalt dynamicity "is a perceptible process of two–sided part–whole dependency in which both parts and whole become perceptually meaningful through mutual interaction that appears as a happening to the whole via its parts. This dynamic interdependency prevents both absolute whole homogeneity as well as whole primacy and absolute part heterogeneity as well as part primacy" (Stadler 2020: 238). Consequently, a Gestalt "involves a back and forth movement that reverses the hierarchy between parts and whole continuously" (Stadler 2020: 286).

[55] Another important claim is that "[s]pecific and definite instances form better gestalts than general or unspecific entities" (Radden and Kövecses 2007: 22).

Relatedly, Barsalou (2017: 3) evokes shape as one set of features distinguished by neural systems in the brain and Gärdenfors (2020: 2–4) claims that an object's shape, or "the relative locations of different parts of an object, can be described in terms of different types of invariances" which "the brain is prepared to pick".

While the universal nature of shape and the ensuing importance of contour-related properties, or boundedness, for characterizing (more or less) prägnant Gestalts is neatly aligned with the overall cognitive (and embodiment) commitment of C(C)G, the impact of embeddedness upon shapes should not be overlooked. In other words, although in general boundedness "is effected by the spatial boundary defining an object's shape" (Langacker 2008: 152)[56], shapes of particular objects are at least partly motivated by their meanings (functions) (Ungerer and Schmid 1996; Tversky 2001; Pinna et al. 2015) and are connected "with the environments in which they exist, with the types of mental set, traditions, habits, with which they are associated" (Smith 1988: 65, after Rausch 1966). An interplay between universality and relativity can also, expectedly, be detected if Gestalt principles are demonstrated at the conceptual level.

3.1.1.2. Construal operations and basic concepts

As noted by Croft and Cruse (2004: 3), in cognitive linguistics, "[p]sychological models of attention and perception, especially Gestalt psychology, have led to the explication of many conceptualization processes in semantics". These meaning-related processes are referred to as construal operations. As Kövecses (2015: 16) puts it, "cognitive operations—commonly called 'construal operations' (...) serve human beings to make meaning—to make sense of their experience, including language".

While it might be possible to provide "a list of construal operations that cognitive linguists typically work with" (Kövecses 2015: 17), e.g. schematization, attention, prominence, perspective, metaphor, metonymy and conceptual integration, a closer look at available sets of construal operations, such as those proposed by Langacker (1987, 2008), Talmy (2000) and Croft and Cruse (2004), reveals that the classifications are only partly overlapping due to, among others, an expanding coverage of construal operations. As Verhagen (2010: 57) puts it,

> [t]he increase in coverage of construal operations in Croft and Cruse's classification in fact goes hand in hand with a (…) decrease of its transparency. It looks as if any new construal

[56] As Langacker (2008: 136) notes, "[d]efined more abstractly, a thing is bounded when there is some limit to the set of constitutive entities".

> operation being discovered requires its own new category. Obviously, this does not mean that certain construal operations must therefore be excluded from the theory, but rather that construal operations may vary in so many different respects that attempts at an exhaustive classification necessarily have a considerable degree of arbitrariness.

Thus, as Verhagen (2010: 57) concludes, it is appropriate to "simply present the most important and well-studied types of construal operations", such as those evoking Gestalt principles.

Gestalt-based construal operations are present in Langacker's, Talmy's and Croft and Cruse's classifications in two ways. Firstly, the Gestalt principle of figure-ground segregation (see Section 3.1.1.1.2.) is seen as a construal operation which not only functions on its own but also underlies other construals, such as the profile-base distinction and the trajector-landmark alignment[57] (Langacker 2008). Secondly, Gestalt principles of grouping and reification (see Section 3.1.1.1.2.1 and 3.1.1.1.2.2.) find their analogies in a number of construal operations pertaining to matter, or THING, to use Langacker's (2008: 105) terminology. THING (Fonteyn 2019), or Gestalt (Paradis 2005), construals encompass, among others, bounding-debounding and melding- discretizing (Talmy 2000)[58]. In other words, matter can be conceptualized on a gradual scale between the extremes of boundedness-unboundedness and homogeneity (continuity)-heterogeneity (discontinuity) (Langacker 2008). Importantly, prototypical instantiations of matter, i.e. physical objects, are maximally bounded and continuous (homogeneous) - "[a] typical object is both continuous and has a definite spatial boundary" (Langacker 2008: 107). Thus, it represents "the special circumstance where grouping and reification are so automatic that constitutive entities are never consciously accessible" (Langacker 2008: 107). To put it differently, a typical object is construed as having prägnant Gestalt properties, i.e. it is maximally bounded and continuous[59].

A focus on conceptual objects and their properties, i.e. products of construal operations, is motivated by the fact that in C(C)G "there is no sharp boundary between simple concepts and certain basic cognitive abilities. We can describe focal red as either

[57] The notion of trajector refers to the primary participant in a relationship, i.e. the entity that is described (the focus of attention) in relation to another object (the landmark). For example, in "The cat is on the mat", "the cat" is the trajector (Langacker 2008).

[58] As emphasized by Croft and Cruse (2004: 63), Gestalt, or constitutive, construals "represent the most basic level of constituting experience and giving it structure or a Gestalt, as described by Gestalt psychologists (Koffka 1935; Wertheimer 1923 [1950]) and phenomenologists such as Husserl (who uses the term 'constitution' in a similar context; see Husserl 1948 [1973])".

[59] Boundedness is used here in the general sense of possessing a contour, i.e. bounded is roughly equivalent to contoured.

a minimal concept or else the ability to perceive this color" (Langacker 2008: 33). Analogically, the ability, or construal operation, of grouping can be described as the simple, or basic, concept of a group. To be more specific, Langacker (2008: 33–34) distinguishes three categories of basic concepts: conceptual archetypes (e.g. a physical object, a whole and its parts and a container and its contents), minimal concepts (e.g. line, angle and curvature), and configurational concepts (e.g. proximity, group, boundary and continuity).

While the configurational concepts of C(C)G are directly related to Gestalt, i.e. a group, and its properties, there are also other categories of Gestalt-based concepts proposed in cognitive linguistics. These include, among others, Talmy's (2000) configurational structures, e.g. state of boundedness, image schemas, i.e. "highly schematic gestalts which capture the structural contours of sensory–motor experience" (Hampe 2005: 1), such as the OBJECT schema and its subsidiaries, e.g. BOUNDEDNESS-UNBOUNDEDNESS and UNITY-MULTIPLICITY (Santibáñez 2002; Szwedek 2019), or schematic ontologies, more neutrally referred to as schematic structures (Paradis 2005: 546), such as THING and BOUNDEDNESS. Thus, in one guise or another, (construal-based) basic concepts embrace Gestalt principles and properties, among which boundedness, i.e. contour possession, seems to occupy a central position[60]. Co-constituted by Gestalt, basic concepts enable the acquisition of a conceptual system, i.e. dynamically organized, categorized knowledge about the world (Kiefer and Barsalou 2013: 4), and form its relatively stable part (Kövecses 2015).

3.1.1.3. The conceptual system

Arising from construal, e.g. Gestalt, operations, alternatively theorized as basic concepts, e.g. Gestalt schematic structures, the conceptual system comprises a network of (more or less) elaborated concepts (see Section 3.1.1.3.1.), which can be represented as a two-level frame (see Section 3.1.1.3.2.).

[60] The term Gestalt schematic structures is adopted here to refer to Gestalt-based conceptual objects, i.e. groupings, and their properties, e.g. boundedness and continuity. This choice is primarily motivated by the fact that although the name schematic structures is already present in construal-related publications (Paradis 2005), it is not as marked as, for instance, image schemas, structural schematizations and configurational concepts. Thus, by introducing Gestalt schematic structures a firm but fairly neutral link is established with existing research.

3.1.1.3.1. Concepts

As already noted above, basic concepts, including Gestalt schematic structures, i.e. groupings and their (more or less prägnant) properties, "are fundamental to the construction of our mental world" since they "represent salient, essentially universal aspects of everyday experience, as determined by the interplay of biological and environmental factors" (Langacker 2010: 3). In fact, basic concepts, i.e. "facets of our common biological make-up" (Langacker 1993: 454), interact with experience in two ways.

Firstly, basic concepts are "triggered or refined by experience" (Langacker 1993: 454) or, to paraphrase Barsalou (2017: 5), "evolutionary-based neural architecture anticipates the space of possible features that could be encountered (…) [e.g. shape], with the experience of specific features then pruning and entrenching the features that remain [in a way] that a language community anticipates will be important for its members". To illustrate, "while in Korean, words for containment [roughly equivalent to boundedness] obligatorily express whether it is tight or loose, as opposed to English in which the distinction is optional", and thus reflect different prunings of the two communities, "both languages express containment" (Mandler and Cánovas 2014: 513), and thus show the ultimate similarity between the two communities.

Secondly, basic concepts are "fleshed out" (Hart 2016: 339) by experience, i.e. the skeletal architecture of basic concepts is enriched by content. For instance, the Gestalt schematic structure of boundedness, i.e. "a schematic template of high generality", can be matched, or imbued, with content structures, i.e. "particular knowledge domains" (Paradis 2005: 546), such as the concrete, e.g. people and artefacts, and abstract, e.g. knowledge and time, domains proposed by Schmid (2000) and utilized by Yilmaz (2020) (see Section 2.2.2.2.2.). Thus, for instance, cars can be described with reference to both Gestalt schematic structures of boundedness and continuity and content structures of vehicle body styles, e.g. coupes, sedans and hatchbacks. A more comprehensive illustration of how schematic and content structures are interwoven is a

> dialectic relationship (…) between a letter of the alphabet and its representation in a variety of different fonts. Although these graphic representations may differ considerably in their fine-grained detail, they can all be recognized as manifestations of the same schematic configuration. In principle we can describe this configuration, thereby defining the letter's 'essence'. We can also describe how each font-specific letter elaborates this schematic characterization. But these two components cannot be dissociated — neither stands alone. Since the defining configuration is quite schematic, it cannot be manifested independently; any actual graphic representation is more specific. Nor can a graphic representation occur without the schema (Langacker 2014: 33).

Importantly the range of content structures elaborating schematic structures can be further divided into intrinsic and extrinsic. A letter, for instance, can be characterized with reference to shape, weight or size (intrinsically) or described with reference to its cultural role, such as being "lucky" or "unlucky" (extrinsically)[61]. As Langacker (1987: 160–161) further explains,

> [s]hape, for example, is a highly intrinsic property of physical objects, as it reduces to relations between the parts of an object and does not require interaction or comparison with other entities. Size, on the other hand, implies comparison either with other objects or with some scale of measurement; hence it is not quite so intrinsic as shape. Behavioral properties tend to be less intrinsic, for most behaviors involve interaction with other entities. Some behaviors are fairly intrinsic, e.g. the sounds that cats emit, and their techniques for washing themselves (…). Such activities as chasing mice and scratching furniture bring external entities into the picture more saliently and are consequently more extrinsic. The cultural role of cats, for instance their association with witchcraft and Halloween, is highly extrinsic; it has little to do with cats themselves, but is rather a matter of how others regard them.

All this knowledge about cats is the content of the concept of cats. To be more specific, "visual information about how cats look becomes integrated in the simulator [i.e. concept], along with auditory information about how they sound, somatosensory information about how they feel, motor programs for interacting with them, emotional responses to experiencing them, and so forth. The result is a distributed system throughout the brain's feature and association areas that accumulates conceptual content for the category" (Barsalou 2005: 624).

Concepts, mediating between embodiment and embeddedness and encompassing a continuum between Gestalt construals and content structures[62] can be represented as two-level configurations composed of "an underlying frame" and "the potentially infinite

[61] In Chinese, for instance, A and S are lucky whereas F and Z are unlucky, which, as Hoon Ang (1997) argues, is related to, e.g., the system of educational grades.

[62] This continuum can be presented in the following way: (EMBODIMENT/BIOLOGY Gestalt (construals (schematic structures (intrinsic and extrinsic content structures EMBEDDEDNESS/SOCIO-CULTURE)))).

set of simulations [conceptualizations][63] that can be constructed from the frame" (Barsalou 1999: 586)[64].

3.1.1.3.2. Frames

As argued by Chen (2001: 203), "[t]o capture the structure of a concept, we must highlight the interconnections among its elements. (...) Concepts should [thus] be represented by such apparatus as frames, which are most effective in illustrating relational information". In other words, "[i]t is not sufficient to characterize a concept as a conjunctive list of binary features (e.g., representing the concept of FACE as eyes & mouth & nose). Instead, considerable research across the cognitive sciences illustrates that concepts contain additional structure associated with arguments and values, conceptual relations, and recursion—concepts are not 'flat' structures" (Barsalou 2017: 10)[65].

Such non-flat structures are anchored in two fundamental ontological categories: events and objects, with the former built on temporal and the latter on spatial relations (Chen 2003: 967). Spatial relations, in turn, rely on "partitioning, in which the mind extends a boundary around a portion of what would otherwise be a continuum of space, and ascribes to the contents within the boundary the property of being a single-unit entity. In such a partitioning process, contents that are perceptually salient, such as those having a clear boundary or those identifiable by shape, would be identified and ascribed quickly

[63] The distinction between concepts and conceptualizations is far from clear-cut. Concepts, for instance, have been used to define "temporary conceptualisations of categories constructed in working memory" (Barsalou 1993: 34), "people's psychological representations of categories" (Barsalou et al. 1993: 1), or "specific mechanisms that represent categories" (Barsalou 2012: 251). The definition adopted here is closest to Barsalou's (2009: 1282) distinction between a simulator and a simulation, where "a concept represents a kind generally [and] a conceptualization provides one specific way of thinking about it" (Barsalou 1999: 587). In other words, "[t]he simulator works as a category type, i.e., a mentally-represented standard on the basis of which simulations of category instances can be constructed. However, the simulator is not an abstract conceptual representation of a concept in the traditional cognitivist sense. Rather, the simulator is a generative neuro-cognitive mechanism that allows us to produce the category simulations" (Langlotz 2010: 176).

[64] While the notion of a frame is not officially employed in a more domain-oriented model of C(C)G, an overall analogy between a frame and a domain is nevertheless acknowledged (Langacker 2014). This analogy derives from a more fundamendal similarity between Langacker's and Barsalou's approaches through which they share "the same basic architecture, involving abstraction from experience, the flexible activation of abstracted symbols in top-down processing, and their recursive combination producing an open-ended array of complex conceptualizations" (Barsalou 1999: 625). Thus, for instance, C(C)G highlights, but not yet fully incorporates into its architecture, recursion, or fractal organization, "whereby analogous configurations occur at multiple levels of organization" (Langacker 2012: 46), i.e. a mechanism typical of Barsalou's frames.

[65] In the same vein, Langacker (2017b: 272) emphasizes that a concept is not "a 'flat' structure such that the nodes are all on the same level, [instead] an assembly includes elements at different levels of organization".

and frequently" (Chen 2003: 965). Object frames are thus evidently constrained by "Gestalt mechanisms of perceptual organization" (Barsalou and Billman 1989: 172–175).

To be more specific, object frames are assemblies of attributes, values, recursion and conceptual relations. Attributes are concepts applied to particular categories. For instance, the Gestalt schematic structure of boundedness becomes the schematic attribute of boundedness when "it describes an aspect of at least some category members" (Barsalou 1992: 30). In other words, "[a] concept is only an attribute when it describes an aspect of a larger whole. When people consider [e.g. boundedness (contour)] in isolation (…), it is not an attribute but is simply a concept" (Barsalou 1992: 30). Attributes, in turn, can be specified into values. For instance, the attribute of boundedness can have the values of polygon/non-polygon. The former, in turn, displays the values of rectangular/non-rectangular, with the latter composed of sidedness and pointedness, each with further values (Pinna 2011). A more specific illustration comes from Szawerna (2012: 204), who claims that "the digital [with a fixed number of clearly defined choices][66] range of values making up the attribute of shape makes it possible to account for comic book covers which are shaped like a vertically oriented rectangle, a horizontally oriented rectangle, a square, etc". In other words, the frame of comics is said to incorporate intrinsic content structures, i.e. shapes of comic book covers, anchored in the Gestalt schematic structure of boundedness (contour). In addition, the frame also includes more extrinsic content structures, i.e. pertaining to comics viewed as cultural rather than material objects. Hence, "the rather specific properties 'Comics narratives are episodic' and 'Comics narratives are open-ended' can be straightforwardly characterized as values of the property 'Comics are dramatic narratives', which in turn becomes an attribute when considered in relation to the two specific properties" (Szawerna 2012: 202).

Situated at different levels of specificity, attributes and values are nevertheless recursively connected into hierarchies. As Szawerna (2012: 201) notes,

> [c]onsequently, a concept's status is a relative matter: while a concept may function as a value of a superordinate attribute, it may also function as an attribute with relation to its subordinate value. Since a concept can be resolved into more specific concepts at any level of abstraction, frames may be embedded in other frames to create hierarchically arranged frame systems.

[66] The other range of values being analogue, i.e. with a continuous range of choices (Szawerna 2012: 204).

In a similar vein, Löbner (2015: 35–36) observes that "[a]ttributes can be bundled into complex attributes; for example, the attributes LENGTH and WIDTH in a rectangle frame can be combined into SIZE. Conversely, attributes with complex values such as COLOR can be split into components". Analogically, as already noted above, the Gestalt schematic structure of boundedness (contour) can be divided into components, i.e. values pertaining to specific contour elements, such as sides or points (corners).

While recursion relates attributes and values into hierarchies, conceptual relations, i.e. structural invariants and conceptual constraints, hold within attributes and values, respectively. To be more specific, "[s]tructural invariants capture relatively constant relations between attributes [while] constraints capture relations between the values of these attributes in specific exemplars" (Barsalou 1993: 40). Importantly, structural invariants "represent relations between attributes that hold true across most exemplars of a category", i.e. between core attributes, e.g. "[a]cross most chairs (…) the seat and back maintain a relatively invariant structural relation to one another" (Barsalou 1993: 39).

All in all, "human conceptual knowledge appears to be frames all the way down" (Barsalou 1992: 40). Still, "only the lesser part of our concepts can be expressed in language" (Löbner 2002: 174), or "indexed by words (...) and used in the service of understanding" (Petruck 2003: 1)[67]. More specifically, Kövecses (2015: 37) claims that

> the frames for concepts are conventionally associated with word forms. In other words, frames are linguistically coded. The use of words for particular frames evokes the frames associated with them. (...) In addition, the words for concepts that are part of a frame also evoke the whole frame. For example, the concept goalkeeper is one of the entities that make up the soccer/football frame. The word goalkeeper thus evokes the entire frame for soccer/football. I do not claim, however, that words (or concepts) always evoke the entire frame that they are associated with or to which they belong. Often, in particular communicative situations several elements from several frames will be assembled to constitute particular mental spaces (Fauconnier, 1985, 1997). I suggest that frames are aspects of long-term memory (...), whereas mental spaces are temporary or online, though structured, assemblages of concepts that emerge in particular communicative situations.

In other words, "linguistic constructions can be modelled as associations of linguistic forms with simulators [i.e. concepts]. On this interpretation, uses of linguistic forms activate (and thus correlate with) the running of simulations [conceptualizations]

[67] This is consistent with the view "that the meanings associated with linguistic units such as words, for example, form only a subset of possible concepts. After all, we have many more thoughts, ideas and feelings than we can conventionally encode in language. For example, as Langacker (1987) observes, we have a concept for the place on our faces below our nose and above our mouth where moustaches go. We must have a concept for this part of the face in order to understand that the hair that grows there is called a moustache. However, there is no English word that conventionally encodes this concept" (Evans et al. 2007: 8).

of the entities or situations to which they refer" (Günther 2016: 100). The ensuing distinction into concept-based constructions and conceptualization-based constructs is foundational to the constructionist commitment of C(C)G[68].

3.1.2. The constructionist commitment of Cognitive (Construction) Grammar

As noted in Section 2.2.1.2., what precisely counts as a construction within a usage-based approach is far from settled. In fact, after 30 years of construction-oriented research, linguists still debate "[o]n what a construction is" (Haspelmath 2023). An overview of this debate is provided by, for instance, Taylor (2012: 124), who rightly concludes that "[t]he term 'construction' has been used in the linguistics literature in a number of ways". While there is "some overlap amongst the different uses, in the sense that there are phenomena which would count as constructions on any of the approaches. Equally, some things which might count as a construction on one of the approaches would be excluded on other understandings of the term". Thus, it is important to precisely define and characterize constructions in C(C)G, which is the focus of Sections 3.1.2.1. and 3.1.2.2., respectively.

3.1.2.1. Defining constructions in C(C)G

Situated within inventory-based construction grammars and thus related to, among others, Goldberg's CxG (Evans and Green 2006: 481), Langacker's approach, expectedly, takes a construction as a basic unit of analysis. In C(C)G "[a] construction is simply an assembly of symbolic structures" (Langacker 2003: 50)[69]. "A symbolic structure [in turn]

[68] This distinction can be elaborated in the following way: "when information about the category is needed (e.g., cued by a word (…)), a very small subset of this accumulated information becomes active to represent it — what might be considered a conceptualization of the category. The entire body of accumulated conceptual knowledge for the category never becomes active all at once — the entire 'concept' is never fully expressed. Instead, only one specific conceptualization of the category is expressed on a given occasion, from the infinitely many conceptualizations possible" (Barsalou 2019: 226). In other words, concepts are never actually realized in toto. Rather, it is only the contextually relevant aspects which surface in (…) use. Borrowing an analogy from phonological theory, we can liken the distinction between (...) concepts on one hand, and their contextualized instantiations on the other as akin to the distinction between phonemes and allophones. Just as with phonemes, (...) concepts qua mental representations [or frames] are never actually perceived. Rather, their existence is inferred based on the variability, and commonalities (...) across (...) many instances of use (Evans 2009: 71).

[69] While Langacker (2003: 50) defines "typical" constructions as composed of two structures (e.g. *deep water*), he at the same time acknowledges the existence of atypical, or degenerate, symbolic assemblies, i.e. such that are composed of just one structure (e.g. *deep, water*). In the current project, the term construction is used in its broadest sense, and includes not only typical but also atypical symbolic assemblies.

reduces to the pairing of a semantic and a phonological structure (its two poles)" (Langacker 2008: 174).

As Langacker (2008: 15) elucidates, "[s]emantic structures are conceptualizations exploited for linguistic purposes, notably as the meanings of expressions". Phonological structures, including sounds and orthographic representations, can be

> overtly manifested [and] hence able to fulfill a symbolizing role[70]. Symbolic structures are not distinct from semantic and phonological structures, but rather incorporate them. (…) The morpheme *cat* can thus be given as [[CAT]/[cat]], where [CAT] stands for the complex conceptualization comprising its semantic pole, and the phonological pole is rendered orthographically in lowercase (Langacker 2008: 15).

Importantly, the notational convention adopted by Langacker, i.e. the square brackets used to render the morpheme "cat" as a symbolic assembly (a construction), means that semantic, phonological and symbolic structures are, at least canonically, well-established or unit-like (see Section 3.1.1.). At the same time, however, it should be underlined that "[s]emantic units are abstracted from the contextual understanding of occurring expressions, phonological units from apprehension of their phonetic properties, and symbolic units from the pairing of the two" (Langacker 2008: 220). Simply put, symbolic units are related to usage events in the way presented in Figure 6 below.

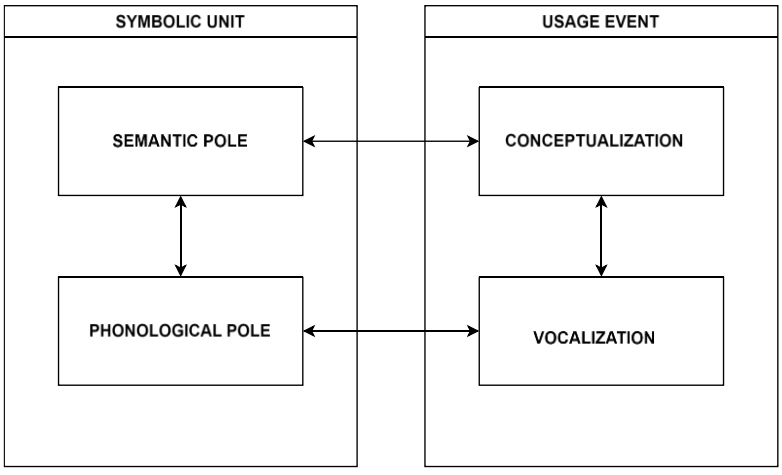


Figure 6. Relations between symbolic units and usage events (adapted from Evans and Green 2006)

[70] Langacker (2008: 15) elaborates on the idea of phonological structures in the following way: "For most linguistic purposes, we are more concerned with the cognitive representation of phonological structures than with their actual physical implementation. It is thus coherent to posit abstract ('schematic') phonological structures which, per se, cannot be overtly manifested". At the same time, however, one cannot fail to notice that phonological structures are far less explored in C(C)G than their semantic counterparts (see Gras and Elvira-García 2021 for a discussion). In other words, while both phonological and semantic structures are roughly equivalent to concepts (Langacker 1987), the former are far less frequently represented in a frame-like fashion than the latter.

Describing the above diagram, Evans and Green (2006: 479), explain that "[t]he horizontal arrows represent (…) correspondences between the conventionalised units of knowledge in the mind of the speaker and the (vocal or conceptual) systems they interact with in instances of situated language use". In other words, the semantic pole is "the multifaceted field of conceptual potential" (Langacker 1987: 76), or "a fixed configuration in semantic space" (Langacker 1991: 290), and a conceptualization is its concrete realization. Likewise, the phonological pole is "our [multifaceted] knowledge of speech sounds" (Langacker 1987: 78), or "an established configuration in phonological space" (Langacker 1991: 290), a vocalization is its concrete realization[71].

The intimate relation between language knowledge and language use, or competence and performance (see Section 1.2.), means that constructions can be approached more or less liberally.

> On the most liberal interpretation (…), any association of a form and a meaning constitutes a construction. The word 'tree' is a construction, since it associates a concept with a phonological form. The sentence you are now reading is, once again, a construction, in that it associates a phonological (or graphological) form with a meaning. A narrower approach restricts the term to form–meaning pairings which have unit status. 'Construction' thereby becomes synonymous with the symbolic unit of Cognitive [Construction] Grammar. On this more restricted understanding, the sentence you are now reading is not a construction, since this specific combination of words and phrases has not become established through usage (Taylor 2012: 125).

While symbolic units may indeed be viewed as constructions par excellence, all symbolic assemblies can in fact be characterized along a cline of constructionhood.

3.1.2.2. Characterizing constructions in C(C)G

As Ungerer (2023: 5) observes, Langacker's C(C)G "can be seen as the archetype of the gradient view" (see Section 2.2.1.2.). To be more specific, "a gradient notion of constructions entails that it cannot be precisely delimited what is a construction and what is not; but in Langacker's [1987: 60] view, this conclusion is 'both acceptable and realistic'". As Langacker (2005: 107–108) notes,

[71] At the same time, Langacker (2008: 46) observes, there are "a lot of terms that might all be applied to the same conceptual phenomenon: concept, conception, conceptualization (…). Although usage varies and the contrasts are subtle", what should be clearly indicated is the choice of terminology and its understanding. In this dissertation, particularly in Chapters 5 and 6, following the already-made commitment to Barsalou's (1999) model of concepts-as-simulators, i.e. two-level frames, the term concept/frame is used to refer to the semantic pole of a construction, while the term conceptualization is employed for the contextualized manifestation of this frame. In other words, my concept/ frame is roughly equivalent to Langacker's (2008) conceptualization and my conceptualization is akin to his conception.

> [s]ymbolic assemblies vary along several dimensions, all matters of degree. First is the extent of their internal symbolic complexity, i.e. the incorporation of smaller symbolic elements. A second dimension is degree of specificity (conversely, schematicity), whether they are characterized in fine-grained or coarse-grained fashion. A third factor is whether they are psychologically entrenched, i.e. mastered as familiar units, and a fourth is the extent of their conventionalization in a speech community. These last two factors determine the degree to which a symbolic assembly constitutes a conventional linguistic unit (whether it is an established element of the language)[72].

A related set of dimensions is proposed by Schmid (2020: 234), who suggests that "the 'likelihood' with which speakers form a schematic representation [i.e. a construction] depends on the frequency, similarity, and paradigmatic variability of a pattern's instances, as well as its syntagmatic size" (Ungerer 2023: 7). Moreover, while "[e]ach of these factors can be quantified using appropriate corpus-based and experimental methods" and "researchers [e.g. Hilpert 2021; Dunn 2022] have [indeed] started to combine several of the aforementioned factors and methods to determine degrees of constructionhood in a data-driven way", there are still uncertainties pertaining to how these factors should be operationalized and "how their interaction should be modeled" (Ungerer 2023: 7–9).

A plausible option seems to involve concentrating on the frequency-based fixedness/novelty cline, i.e. "a first plausible determinant of constructionhood" (Ungerer 2023: 7) (Section 3.1.2.2.1.), with specificity/schematicity (Section 3.1.2.2.2.) and simplicity/complexity (Section 3.1.2.2.3.) providing further details about the (already-) established constructions.

3.1.2.2.1. The fixedness-novelty continuum

Since "symbolic assemblies vary in the extent to which they achieve the status of units and become conventional within a speech community" (Langacker 2008: 21), the fixedness-novelty continuum can be observed at two levels: individual and communal.

At the individual level, fixedness, i.e. unithood, is related to entrenchment. While the phenomenon is "extremely complex" (Blumenthal-Dramé 2012: 187), "[t]he first force on entrenchment that comes to mind is, of course, frequency of repetition". Simply put, "the more frequently a given word or structure has been processed, the more deeply the patterns of associations that are activated for processing willbecome entrenched.

[72] In other words, it seems that Langacker employs the following terminology: entrenched (established at an individual level), conventionalized (established at a communal level), and conventional (established in a language). In this dissertation, conventionalized/conventional are used interchangeably to refer to stability at a communal level and the term fixedness is used to refer to symbolic assemblies established at either or both levels.

Therefore, frequency of rehearsal and repetition of whatever we can identify as being similar in some respect (...) are expected to contribute to the learning and consolidation processes" (Schmid 2020: 216).

Importantly, however, even though "entrenchment is essentially an experience-based learning process" (Schmid 2020: 206), there are also patterns which "become so important over evolution that the brain evolves to represent them with minimal learning" (Barsalou et al. 2003: 75). Relatedly, Deane (1993: 194) observes that "[t]ime of acquisition is (...) critical to entrenchment. A concept which is acquired at age one is almost certain to be more entrenched than a concept which is only acquired at age ten". Such "developmentally early concepts are likely to be of the sensorimotor variety, concerned primarily with bodily interaction with the immediate environment" (Deane 1993: 195). Likewise, Schmid (2020) claims that embodied processes and their products function as the foundation for comprehending the environment since they are either innate or acquired very early. As Schmid (2020: 222) claims,

> this "ontologically early and deep cognitive entrenchment" (Schmid 2007: 125) includes, among others, "the domain of space, possibly represented in the form of image schemas such as up–down, behind–in front, left–right, in–out, source–goal, centre–periphery (Bergen 2015, Johnson 1987, Lakoff 1987); basic principles of perception and attention, e.g. figure-ground segregation and other Gestalt principles (Ungerer and Schmid 2006: 34–44, 163–166).

All in all, then, while recurrence is central for the entrenchment of constructions within individual minds, entrenchment is at the same time mediated by other factors, e.g. embodiment, and thus certain constructions, particularly those whose semantic pole is "based on construal (…) and basic cognitive abilities", e.g. grouping and reification (Langacker 2014: 39), seem more adequately described as triggered or pruned by experience (see Section 3.1.1.3.1.) rather than learned.

Moreover, since learning, and thus entrenchment, takes place "through interaction in a physical, social, cultural, and discourse context (Langacker 2014: 30), entrenchment involves the knowledge of conventions. As Langacker (2008: 30) elucidates, "[a]n individual's notion of what an expression means develops through communicative interaction and includes an assessment of its degree of conventionality in the speech community". What might enter an individual's long-term memory is thus, for instance, the sociolinguistic fact that "the symbolic unit 'ain't' is stigmatized, that 'sir' indicates the relative social status of speaker and addressee, that 'déjà vu' is a borrowing from French" *(*Langacker 1987: 62). As Schmid (2003: 8) explains, "people do not accumulate and structure their long-term memories from their own individual resources but under the

influence of the cultures they are exposed to", or the cultures in which they are embedded or have access to. One example of how individual resources are shaped by those of other people is when "the meaning of a category has to be learned thanks to the contribution of other members of our community", i.e. when "[o]ther people can help us understand (...) concepts providing us with explanations, or furnishing us a list of possible instances of the category. When hearing or reading new terms, we often search for their meaning in the dictionary, or look up their meaning on Wikipedia" (Borghi and Binkofski 2014: 30).

All in all, individual and communal levels "interact with each other in an intricate manner. On the one hand, the individual cognitive environments of the members of a culture are shaped by the collective knowledge, beliefs, traditions etc. On the other hand, cultural environment can be considered as the collective long-term memory of a group" (Schmid 2003: 8).

The collective long-term memory of a group, i.e. its shared knowledge and beliefs, is the seat of conventionalized symbolic assemblies, i.e. constructions fixed at the communal level. As Langacker (1987: 62) explains, "[c]onventionality implies that something is shared—and further, that it is recognized as being shared—by a substantial number of individuals". Thus, like entrenchment, conventionality, or conventionalization, is also closely related to frequency[73].

In fact, Schmid (2016: 551) observes that "[f]rom a cognitive-linguistic perspective [including C(C)G], (…) increasing diffusion [and thus conventionality] simply corresponds to more distributed entrenchment, i.e. to an increase in the number of members of a community who have a representation". At the same time, though, Schmid (2016) notes that representations, or symbolic assemblies, are entrenched to different degrees by different members of a community. Thus, discovering stabilities at the communal level, e.g. in a corpus[74], is not tantamount to revealing (equal) entrenchment

[73] While conventionality as "[t]he objectively given distribution of units across a speec community" (Langacker 1987: 62) can be operationalized in terms of frequency, the way in which a community realized that it shares a construction is harder to grasp. One plausible option seems to operationalize this collective awareness as the existence of explicit norms, including "linguistic reference works such as dictionaries, grammars, and usage manuals, (…) complemented and to some extent replaced and superseded by online platforms run by publishing houses or crowdsourced with the help of users" (Schmid 2020: 99).

[74] Measuring conventionality through a corpus involves the corpus-as-output perspective, where "a corpus is a sample of the language use of a particular group of speakers representative for a particular speech community—a snapshot, as it were, of the linguistic performance they collectively generate on the basis of the individual linguistic representations in their minds. The corpus-as-output view is not incompatible with a cognitive approach: As in other methodological frameworks relying on the observation of naturally occurring behavior, we can draw inferences about the mental representations underlying this behavior" (Stefanowitsch and Flach 2016: 102–103).

within each of the individual minds constituting the aggregate. What is shown, instead, is "a preference" for a representation, or a construction, exhibited by individuals within a community (Hoffmann 2021: 85–88)[75].

In other words, since entrenchment "describes established linguistic structures at the level of an average speaker's mental representation or at the level of the speech community as an aggregate of individuals", individual entrenchment should be distinguished from "collective entrenchment" (Enghels and Camarero 2021), or conventionality in Langacker's (1987, 2008) terms.

Importantly, however, even though conventionalization could be operationalized in terms of frequency with which a symbolic assembly is employed by members of a particular community, i.e. in terms of a collective preference, it should simultaneously be noted that a given community is rarely, if ever, homogeneous, overlapping, as it were with a number of other groups. As Croft (2009: 12) puts it, "[s]ome communities are quite specialized, such as the community of cognitive linguists. But other communities are very broad and even all-encompassing (…) [e.g.] the community of Western urban industrialized society, the community of men, or ultimately the community of human beings on Earth, the community to which we all belong". Clearly, then, communities are embedded within other communities, e.g. cognitive linguists are a part of the community of linguists and the community of human beings, just as academic ELF users are a part of the academic community, the community of users of ELF (including native and non-native speakers), and ultimately also the community of human beings (see Section 1.2.2.2.). Thus, while analyzing the frequency of symbolic assemblies in a corpus of academic ELF, the focus could be on constructions preferred by, for instance, the ELF academic community as a whole or by one of its sub-communities, e.g. Polish users of academic ELF.

Acknowledging that "[t]he relation between frequency and degree of conventionalization is not that straightforward", i.e. that "speech communities partially overlap, and individual speakers can rather freely engage in multiple different speech communities", Bloom (2021: 20–21) emphasizes that it is "important to remain

[75] A similarly indirect link between entrenchment and conventionalization is advocated by Vetchinnikova (2017: 282), who sees "the processes of fixing at the individual plane and conventionalisation at the communal plane (...) as similar, even though (...) the mechanisms underlying them are different. It seems that this similarity across different planes can be regarded as an example of scale-free self-similarity or fractal scaling (Mandelbrot 1982; Gleick 1987)".

conservative in making generalizations about the overarching speech community, as the observed conventions may be established in a smaller speech community only". Thus, it is advisable to limit one's investigations of conventionality to a clearly demarcated and precisely defined (sub-)community[76].

All in all, whether at the individual or at the communal level, fixedness means that a symbolic assembly has evolved into a relatively stable preference and can thus be contrasted with a novel symbolic pairing, i.e. "a fleeting, one-time occurrence", like the appearance of the verb "sneeze" in the caused-motion construction (Langacker 2009: 153). Seen as a novel usage, and thus as lacking (full) compatibility with the semantic pole of the caused-motion construction, "sneeze" can, nevertheless, develop from a one-time, or emergent, occurrence to a fixed, or emerged, construction[77]. To illustrate this process, Langacker (2009) presents a potential scenario according to which teenagers start using "sneeze" in the caused-motion construction, initially as a fad and gradually as a habit. In this way, "sneeze" "apprehended as a caused-motion predicate, assumes the status of a fixed, familiar unit". While "[t]his has not yet happened for 'sneeze', at least in the general populace. It may be happening in the smaller speech community of linguists, particularly cognitive linguists" who discuss the caused-motion use of "sneeze" so frequently that it is "quite familiar", i.e. no longer novel (Langacker 2009: 153–154).

Clearly, then, there is no definite boundary between fixed and novel symbolic assemblies. What there is instead is a gradation from novel through incipient to established units (Langacker 2008: 38), i.e. a cline from emergent through emerging to emerged constructions (see Section 1.2.2. and Section 2.2.1.2.)[78]. In fact, an interplay between stability and temporality is "typically the case, most expressions are some combination of units and non-units" (Langacker 2017a: 11). For instance, in the case of the phrase "tall penguin", which Langacker (2017a: 11) categorizes as novel, "'[p]enguin' is a unit, 'tall' is a unit, the pattern that combines them is a unit. But the composite expression 'tall penguin' is novel, it's not a unit, it's a new structure".

In other words, classifying a symbolic assembly as tending towards fixedness or novelty depends on the level of granularity adopted. To be more specific, "tall penguin"

[76] Admittedly, this approach is adopted in the study described in Chapter 6.
[77] In the current project, particularly in Chapters 5 and 6, symbolic assemblies are referred to as constructions and classified as emergent, emerging and emerged.
[78] The direction of the proposed cline does not mean that emerged structures are static. On the contrary, they are constantly affected by use and thus reinforced by elaboration, augmented or violated by extension and suppressed by non-use (Langacker 2017a).

is novel if the standard of comparison is a specific adjective-noun combination (the tall penguin construction) but conventional if the point of reference is a schematic adjective-noun pattern (the adjective-noun construction). To put it differently, "tall penguin" is novel, or emergent, since it extends the semantic poles of the two specific constructions, i.e. "tall" and "penguin". At the same time, though, "tall penguin" is fixed, or emerged, since it elaborates the semantic pole of the schematic adjective-noun construction. A detailed account of the specificity-schematicity continuum is provided below.

3.1.2.2.2. The specificity-schematicity continuum

As stated by Langacker (2003: 43), "[b]oth specific expressions and abstracted schemas are capable of being entrenched psychologically and conventionalized in a speech community, in which case they constitute established linguistic units. Specific expressions with the status of units are traditionally recognized as lexical items. More schematic units correspond to what is traditionally regarded as grammar" (Langacker 2003: 43).

While both specific and schematic units are acknowledged in C(C)G, the former enjoy a more privileged position than the latter. To elucidate, while C(C)G does not share reservations expressed by some construction grammarians, e.g. Croft (2001), "about the existence of high-level syntactic generalizations such as the noun phrase, subject and object, or even part-of-speech categories such as noun or verb", and does not support arguments that schematic constructions are "nothing more than a vague idea entertained by speakers who are analytically minded enough to see similarities between different kinds of construction" (Hilpert 2014: 68–69), claiming instead that researchers should not only "be open to the possibility of schematic definitions, but arguably their existence ought to be the default expectation" (Langacker 2009: 174), C(C)G nevertheless views schematic units as less numerous. As Langacker (2010: 50) explains,

> [s]ince language is learned through usage, it ought not be surprising that the preponderance of linguistic knowledge consists of specific expressions and low-level patterns, many of which incorporate particular lexical items. This is not to deny the existence and importance of general, productive patterns represented by high-level schemas. I would however suggest that fully general patterns constitute a distinct minority, that lower-level structures provide critical information and do much if not most of the work in speaking and understanding.

This work entails functioning as categorizing structures for usage events, which, in most cases, is the role performed by specific units. In other words, "more specific units have a built-in advantage over more schematic ones in the competition to be selected as

categorizing structure. Being characterized in finer-grained detail, a specific unit has more properties that might overlap with the target to boost its level of activation" (Langacker 2008: 230). Frequently activated, specific units become more fixed, i.e. their degree of constructionhood increases.

Simultaneously, though, two caveats need to be noted. Firstly, as already mentioned in Section 3.1.2.2.1., certain schematic units, particularly those classified as innate or early acquired, e.g. Gestalt schematic structures, are impervious to frequency effects and thus remain defaultly fixed and accesible. Secondly, "[p]articular high-level schemas may be so entrenched and accessible (relative to alternative patterns or to subschemas) that they normally win the competition and prevail as general or default-case patterns" (Langacker 2010: 50).

Along similar lines, Höder (2021: 41) observes that

> [t]here is an ongoing debate in construction grammar on how much generalization is cognitively plausible – as of today, most constructionists would probably agree that highly schematic constructions are less likely to play a major role if they don't serve any additional purpose beyond what can be achieved by less schematic ones. But sometimes they do play a role, in particular when they're used productively in a way that wouldn't be possible using less schematic constructions alone", i.e. when they are employed for multilingual, including ELF, interactions.

In such situations[79], it seems, schematic constructions tend to be readily activated (see Section 1.2.2.2.) and thus their constructionhood becomes relatively high[80].

Regardless of their frequency of activation, specific and schematic constructions can be felicitously characterized as form-meaning pairings. To begin with, "[a] typical lexical item [i.e. a specific construction] has a specific phonological shape as well as detailed semantic content", encompassing both intrinsic and extrinsic content structures as well as providing access to (Gestalt) schematic structures (see Section 3.1.1.3.1.). In other words, canonically, the semantic pole of specific constructions rests primarily on rich knowledge structures and only secondarily on skeletal ones. Thus, for instance, "glass" and "wine glass" are both specific constructions – lexical units with a particular form and an elaborate meaning (Langacker 2008: 49). Also specific, but to a lesser degree, is the noun "thing", rendered as [[THING]/[thing]], which, while still displaying a concrete graphological form is "schematic semantically" (Langacker 2008: 23). Also semantically schematic is the noun class description, i.e. the [[THING]/[…]] symbolic

[79] A similar account is also offered by Strugielska and Piątkowska (2017).

[80] On the whole, though, a schematic unit's constructionhood is viewed as low "because speakers perceive its instances as too dissimilar to generalize over them" (Ungerer 2023: 14).

unit. However, unlike the noun "thing", the noun class schema is also graphologically non-specific. Thus, as Langacker (2008) contends, specific and schematic constructions form a continuum, which can be viewed at two intertwined levels. Firstly, on a more local level, there is a cline within both specific lexical items and schematic grammatical templates. The former can be illustrated by the following hierarchy: "thing→object→vehicle→truck→pick-up truck" (Langacker 2003: 47), while the latter by the "noun→count noun→common count noun continuum", both of which reflect recursive attribute-value sets discussed in Section 3.1.1.3.2.

Secondly, there is an overall cline between lexical items and grammatical units, i.e. the lexicon-syntax continuum (Langacker 1987), with the former specifying the latter by providing experience-derived values on biologically-motivated attributes, i.e. construal-based basic concepts (see Section 3.1.1.2. for details). Importantly, "as one moves from the lexical end of the continuum toward the grammatical end, (…) the role of construal is more predominant", i.e. the semantic pole of grammatical constructions rests primarily on (Gestalt) schematic structures, which are only to some extent modifiable by experiential details (recall the notions of cognitive styles in Section 3.1.1.1.2. and collective prunings in Section 3.1.1.3.1.). Derived from embodied cognition (see Section 3.1.1.1.1.), (Gestalt) schematic structures characterize "fundamental grammatical notions", a prime example being the noun category, whose semantic pole – THING – is "defined abstractly as any product of conceptual grouping and reification" (Langacker 2014: 39) (see Section 3.1.1.1.2.), and thus describable in terms of, e.g., boundedness.

While boundedness has so far been roughly defined as contour possession (see Section 3.1.1.2), it is now appropriate to clarify that being bounded involves reaching the limit of an actual instance, i.e.

> [a] thing is bounded if (…) the requisite set of entities is eventually exhausted. The instance conception is then complete, in the sense that further scanning through constitutive entities amounts to conceptualizing another instance of the same type. In short, there is some notion of reaching the limits of a single instance, making it possible to begin the conception of another, distinct instance (Langacker 2008: 136–137).

Developed primarily for characterizing nouns, the semantic pole based in grouping and reification is by no means limited to this class. In fact, Gestalt schematic structures, e.g. boundedness, have been felicitously employed for characterizing the semantic poles of, among others, verbs (Talmy 2000), adjectives (Paradis 2001) and quantifiers (Langacker 2016b; Davidse et al. 2017), and further refined as a result. In the case of boundedness

applied to quantifiers, for instance, being bounded involves not so much reaching the limit of an actual instance as reaching a complete coincidence between the limit of an actual instance and a reference, or a virtual, instance. The two interpretations of boundedness, i.e. bounded-as-limited and bounded-as-complete[81], characteristic of countable nouns and relative quantifiers, respectively (Radden and Dirven 2007), are illustrated below.

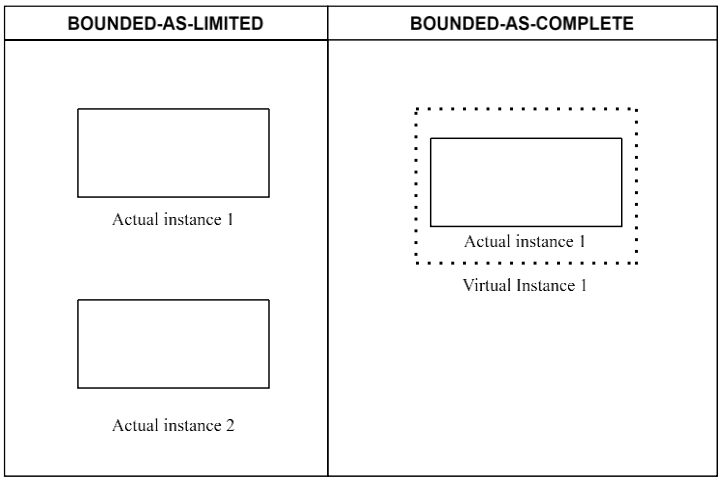


Figure 7. Bounded-as-limited and bounded-as-complete (adapted from Radden and Dirven 2007)

In fact, as the above illustrations imply and as Langacker (2001: 154) explicitly acknowledges, grouping (and, at least some, reification) characterize not only "classic", i.e. linguistic, schematic units but also paralinguistic ones (Cienki 2015)[82], e.g. attentional frames.

Inspired by Chafe's (1980, 1987a, 1987b, 1994) notion of an intonation, or a prosodic[83], unit, Langacker (2001: 155) explains that "at the phonological pole each [attentional frame] consists of some kind of intonational grouping; and at the semantic pole, each comprises a body of active information all of which is simultaneously available". Importantly, attentional frames are symbolic, "in that the very act of imposing a particular intonational grouping effects and symbolizes the act of imposing the corresponding conceptual grouping". In other words, despite their abstract, or highly

[81] From now on, boundedness is used in the sense of possessing an actual limit and completeness in the sense of possessing a virtual limit.
[82] As Cienki (2015: 511) explains, "it is not denied that 'traditional' words and grammar form the expressive core. However, other behaviors can gain symbolic status as well, e.g., non-lexical sounds, intonation contours, and (manual) gestures. (…) That is, in the model of grammar as abstracted from spoken [and, by analogy, written] language usage events, the borderline between the linguistic and the paralinguistic is considered permeable".
[83] As Chafe himself (1987a: 6) notes "[s]poken language exhibits important prosodic units of a kind that [he has] been calling 'intonation units'".

schematic nature, attentional frames, are constructions whose semantic and phonological specifications are different from that characterizing "classic" symbolic assemblies, e.g. lexical and grammatical constructions (Langacker 2001: 155–156). An attentional frame is depicted in Figure 8 below.

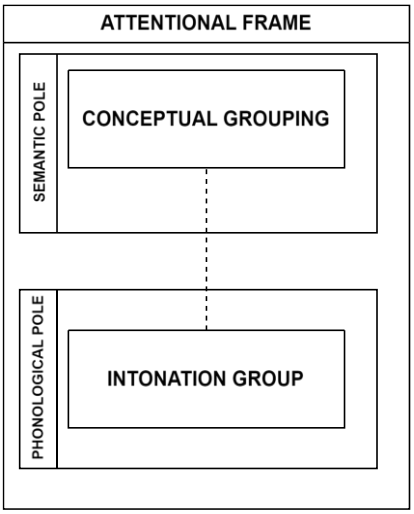


Figure 8. Attentional frame (adapted from Langacker 2001)

While attentional frames are clearly focused on grouping, reification, and consequently further Gestalt schematic structures, can be also discerned. For instance, describing attentional frames, Langacker (2001: 154) refers to single and multiple units, describes them as involving pauses or breaks, mentions terminal contours and changes in quality. In other words, it seems that attentional frames could be felicitously described with reference to at least such Gestalt schematic structures as simplicity, continuity, boundedness/completeness and homogeneity.

Importantly, attentional frames, roughly equivalent to paralinguistic units (see Section 1.2.2.), interact with linguistic units to form intonation/prosodic units (henceforth referred to as prosodic units) in the spoken mode[84]. In the written mode[85], in turn, punctuation units, i.e. "the stretches of language between two punctuation marks" (Chafe

[84] While Langacker (2001, 2008) refers to channels rather than modes, the two are, nevertheless, intimately connected. The spoken mode, for instance, encompasses "the core channel", i.e. "what is often referred to as segmental content: the series of sounds or phonemes an utterance is composed of. A second channel is prosody, consisting of 'suprasegmental' phenomena such as accent, tone, rhythmic grouping, and intonation contours. (…) Less commonly recognized as part of language is the gestural channel, including manual gestures, facial expression, and bodily posture" (Langacker 2008: 461–462).
[85] Modes are discussed in more detail in Section 4.2.

1987a: 3), can appear at different levels and thus encompass, e.g., (parts of) sentences as well as paragraphs[86], i.e. structures of varying complexity.

3.1.2.2.3. The simplicity-complexity continuum

As aptly condensed by Stefanowitsch and Flach (2016: 105), "[t]he dimension of complexity concerns the internal structure of linguistic units, that is, the question of whether, and to what degree, a unit consists of identifiable subunits at the same level of articulation"[87]. At the morphological level, for instance,

> [m]onomorphemic words are maximally simple: They cannot be analyzed[88] into smaller meaningful units. Multimorphemic words are slightly more complex, at least from the perspective of the analyst: They consist of at least one root and one or more affixes. Still further up the dimension of complexity, we find multiword expressions, such as adjective–noun compounds (e.g., 'working class', 'higher education') and fixed phrases (e.g., 'a great time', 'a great deal', 'for old time's sake') (Stefanowitsch and Flach 2016: 106).

To be more specific, "[a] morpheme is definable as an expression whose symbolic complexity is zero, i.e. it is not at all analyzable into smaller symbolic components. A morpheme can also be thought of as a degenerate symbolic assembly [see Section 3.1.2.1.] comprising just a single symbolic relationship" (Langacker 2008: 16). To elucidate the difference between simple, or degenerate, assemblies, and complex constructions, Langacker (2008: 60) emphasizes that symbolic complexity involves

> being assembled out of smaller symbolic elements (…). For example, 'lipstick' has 'lip' and 'stick' as symbolic components. These are component symbolic structures, 'lipstick' as an integrated whole being the composite symbolic structure. Likewise, 'make' and '-er' are symbolic components of the composite expression 'maker'. A composite structure can itself function as a component structure in an expression of greater symbolic complexity. 'Lipstick' and 'maker' are thus components of the higher-level composite structure 'lipstick maker'. Linguists refer to this hierarchical arrangement as constituency and represent it in tree-like diagrams.

Being tree-like, "an assembly is not a [classic] tree structure. One reason is that the same elements are grouped simultaneously in alternative ways based on different functions. For

[86] In a similar vein, Dahl (2018: 26), following Lennard (2005: 151), defines punctuation marks as "non-alphabetic and non-numeral 'marks, spaces and other signs (such as distinguishing type-faces or fonts) placed within the text to articulate, dis/ambiguate, or otherwise refine and/or display the sense'. Apart from segmental marks like the period, comma, etc., as well as spaces between words and paragraphs, Lennard's definition includes various sorts of case variation". In fact, in its broadest rendition, Lennard's (2000: 5–6) definition of punctuation includes "everything between the letter-form and the book itself" (Dahl 2018: 32).

[87] Morphemes, for instance, which typically include more than one phoneme are simple at the level of morphology but complex at the level of phonology.

[88] In Langacker's (2008: 61) words, analyzability of a complex (or composite) structure shows "how salient the component structures are in relation to the composite conception, and how strongly they contribute to its emergence".

instance, the segments of albums are grouped as ((al)(bums)) on purely phonological grounds, and as ((album)(s)) for symbolic purposes" (Langacker 2017b: 273).

These dynamic tree-like assemblies are not unlike attribute-value hierarchies discussed in Section 3.1.1.3.2., particularly that both Langacker and Barsalou consider relations between higher- and lower-level structures, i.e. compositional paths and structural invariants[89], respectively, as vital to meaning. Incorporating relations into the architecture of a symbolic unit

> just seems right—this is a very real dimension of conceptual organization, and to ignore it is simply pointless. More significantly, it helps explain the commonplace observation that no two expressions are exactly the same in meaning. A classic example is 'pork' vs. 'pig meat'. Suppose, for sake of argument, that their composite semantic structures are taken as being identical. They are nevertheless semantically nonequivalent, since they arrive at this composite conception via different compositional paths: a degenerate path in the case of 'pork' (consisting of just the composite structure, since there are no individually symbolized components); and for 'pig meat', a path incorporating both PIG and MEAT. The semantic difference, then, is that pig meat evokes the component notions PIG and MEAT more saliently than does 'pork' (Langacker 2008: 61–62)[90].

Alternatively, complexity and simplicity could be interpreted in terms of elaboration or lack of thereof, i.e. baseline. As Langacker (2016c: 1–2) explains,

> [t]he notions baseline and elaboration pertain to asymmetries observable in any facet of language structure or its conceptual and phonological basis. (…) Taken in the broadest sense, B/E organization is utterly ubiquitous. Figure vs. ground can be characterized in this fashion: when something contrasts with its surroundings, the situation is more elaborate than the baseline situation of uniformity. Any kind of norm can be thought of as a baseline with respect to which deviations constitute elaborations. (…) Linguistic applications of these notions thus include such general matters as silence vs. speech, [and] assessments of normativity (conventionality, well-formedness).

A more specific example of B/E organization is the way in which the head noun, i.e. the baseline, can be elaborated into a noun phrase. As Langacker (2017b: 305–306) states, "[i]n English, the head noun has two dimensions of elaboration: intrinsic and extrinsic. Extrinsic elaboration is done by modifiers. (…) Intrinsic elaboration pertains to the head

[89] The other conceptual relation evoked in this dissertation is that of conceptual constraints (see Section 3.1.1.3.2. for details).

[90] While compositionality is closely related to analyzability (for details of this relation see, for instance, Kardela 2012 and Żyśko 2017), i.e. analyzable (complex) expressions can often be thought of as being assembled from their components" in accordance with regular compositional patterns" (Langacker 1999: 62), Langacker (2008: 170) cautions that "[a]t the extreme, there may be no connection at all between component and composite meanings. The meanings of 'under' and 'stand', for example, play no apparent role in the composite meaning of understand. Though morphologically complex, this verb is semantically unanalyzable" (for a related discussion on the challenges related to complexity/analyzability/compositionality see Ungerer and Hartmann 2023).

noun's internal structure: it creates the basic type. It does so through compounding and morphological derivation (rather than modifiers)". In other words, the structure established through intrinsic elaboration functions as the baseline for extrinsic elaboration. At the same time, though the baseline, e.g. a compound, "exhibits B/E layering. The first element is a reference point: a baseline for interpreting the second element, which imposes its profile on the composite whole" (Langacker 2017b: 305–306). Within this composite whole, the ultimate baseline is the second element, e.g. "meat" in the "pig meat" assembly[91].

While B/E organization is seen as highly dynamic, it "doesn't always manifest itself in strict temporal sequencing. You have to interpret it a bit more abstractly than that", i.e. "as A/D organization, [where] A is autonomous—with the potential to be manifested independently—and D is dependent on A, which it requires for its full manifestation" (Langacker 2017a: 243). A/D organization can be neatly illustrated with reference to prosodic/punctuation units, units, i.e. complex constructions incorporating attentional frames (see Section 3.1.2.2.2). Expanding on the representation of the attentional frame, depicted in Figure 8 above, where only one component of a prosodic/punctuation unit is shown, i.e. the paralinguistic construction, Figure 9 below presents a fully-fledged prosodic/punctuation unit, i.e. a complex construction combining a paralinguistic and a linguistic unit[92].

[91] This logic underlies the architecture of the elaborated abstract construction presented in Chapter 5. To be more specific, the abstract label construction functions as the reference point through which its dominion, with the abstract content construction as the target (Langacker 2008: 84–85), is interpreted. Thus, the abstract label elaborates the abstract content, which is the ultimate baseline in the assembly.

[92] For the sake of clarity, the terms linguistic and paralinguistic are used here as simplified counterparts to what Langacker (2001: 156) refers to as structural structures, i.e. assemblies of segmental and conceptual content, and attentional structures, assemblies linking intonation/prosody (and, analogically, punctuation) and information structure (defined here in terms of grouping (Section 3.1.1.1.2.1.) and reification (Section 3.1.1.1.2.2.).

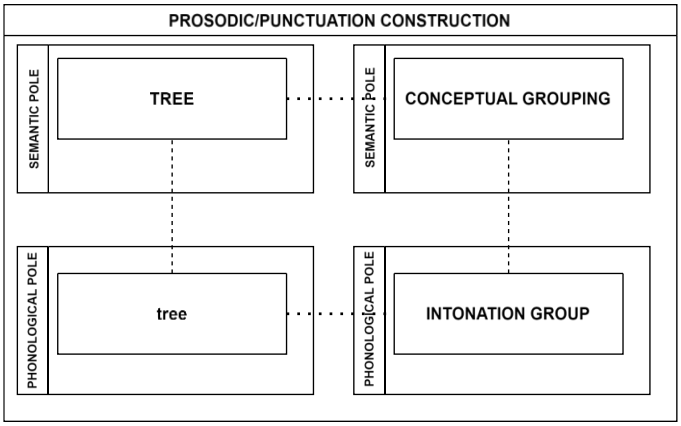


Figure 9. Prosodic/punctuation construction (adapted from Langacker 2001)

Importantly, the complexity present in Figure 9 above can be rendered in a more detailed way through alternative groupings[93], involving dissimilar degrees of analyzability and different compositional paths, and leading to alternative structural invariants and conceptual constraints (see Figure 10 below).

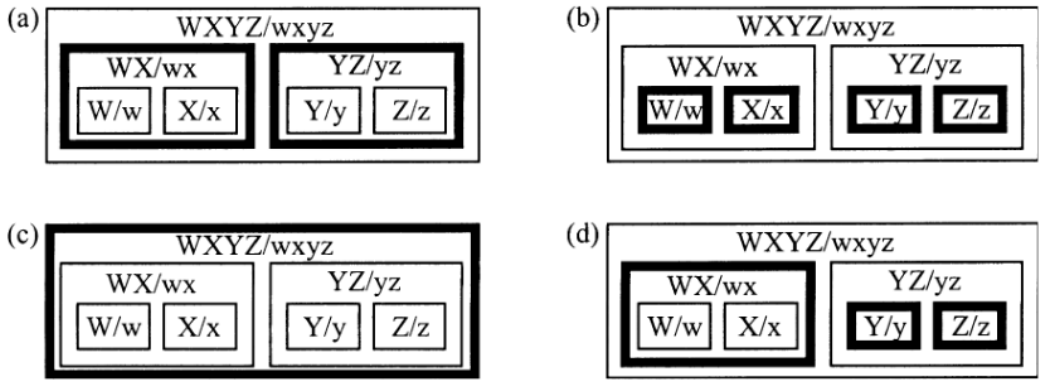


Figure 10. Prosodic/punctuation constructions (Langacker 2001: 158)

3.2. Concluding remarks

The aim of Chapter 3 was to present Cognitive (Construction) Grammar as a usage-based approach with the potential to compensate for the gaps identified in CxG-oriented studies pertaining to academic ELF. Thus, the presentation of C(C)G was guided by the need to (further) highlight links between cognitive processes at individual and collective levels, low- and high-frequency patterns, emerged, emerging and emergent constructions, embodied and socio-cultural dimensions of meaning as well as linguistic and

[93] In other words, a clause can be presented within one punctuation unit, i.e. presenting only a higher-level structure, or additionally showing words constituting a clause within separate punctuation units, i.e. presenting also lower-level structures which are combined to create a higher-level structure.

paralinguistic forms through which meaning could be conveyed. In other words, the possibilites provided by the gradient view upon constructionhood, which were only partly exploited with reference to academic, particularly written, ELF in Chapter 2, were elaborated and expanded in Chapter 3 so that ELF's nature could be interpreted in a more comprehensive manner.

To be more specific, Section 3.1.1. provided a fully-fledged perspective on embodiment and, relatedly, on Gestalt processes of grouping and reification, which are universally available and thus also accessible to ELF (sub-)communities for pruning and collective entrenchment. Section 3.1.2., in turn, showed how Gestalt-based construal operations, a.k.a. Gestalt schematic structures, found the architecture of linguistic constructions. Defined as symbolic assemblies, constructions were characterized as varyingly fixed, specific, and simple. . Concurrently, an interplay between linguistic and paralinguistic units was emphasized and the inherently complex nature of prosodic/punctuation units was duly highlighted. This intricacy will be further explored in Chapter 4, where the focus is on discourse and discourse genres, i.e. par excellence seats of complexity and diversity.

Chapter 4

A Cognitive (Construction) Grammar approach to discourse genres

4.0. Outline

The purpose of Chapter 4 is to present a Cognitive (Construction) Grammar approach to discourse genres. The chapter is divided into two main sections, characterizing discourse genres as constructions (Section 4.1.) and researching discourse genres as constructions (Section 4.2.).

Section 4.1. describes the rationale for attributing the constructional status to discourse genres. The section commences with an explanation of the pivotal role of discourse within the C(C)G framework and further discusses the foundational assumptions related to theorizing discourse genres as constructions.

Section 4.2. illustrates how the theoretical considerations of C(C)G are empirically elaborated and extended. The studies selected for this purpose present particular solutions employed in investigating discourse genre constructions in spoken and written modes, further divided into non-academic and academic discourse genre constructions, thus providing an overview of methodological approaches currently embraced within the aforementioned fields.

The chapter concludes by recapitulating the key premises presented in both sections and highlighting possible avenues for advancing a C(C)G approach to academic (ELF) written genres.

4.1. Characterizing discourse genres as constructions

As noted by Langacker (2008), considerations developed within the framework of C(C)G are not limited to the description of constructions within the boundaries of a single sentence. To be more specific, the "all-inclusive" nature of the model is evident in its commitment to embracing units of varying complexity (Section 3.1.2.2.3.). As Langacker (2008: 457) elucidates,

> [s]tarting from single words, like nouns and verbs, we have worked our way up to successively larger expressions: to multiword constructions, to full nominals and clauses, and finally to complex sentences. The next level is discourse, where any number of sentences (or

> fragments thereof) are connected to form a coherent linguistic production—be it a conversation, a monolog (e.g. a speech), or a written text.

Admittedly, although constructionist approaches have "only recently started to extend to the analysis of larger-than-the-sentence patterns" (Antonopoulou and Nikiforidou 2011: 2594), Langacker (2008) emphasizes that the inclusion of the discourse level in the constructional framework is crucial for a more comprehensive understanding of linguistic knowledge.

4.1.1. Discourse

In Cognitive (Construction) Grammar, discourse is seen as a set of interrelated usage events of no pre-defined size, e.g. sentences, paragraphs, or conversational turns, constituting the current discourse space[94]. As Langacker (2008: 59) explains,

> [a]s discourse unfolds, at each step the current expression is constructed and interpreted against the background of those that have gone before. The prior discourse is a major determinant (along with context, background knowledge, etc.) of what I call the current discourse space (CDS). The CDS is a mental space comprising everything presumed to be shared by the speaker and hearer as the basis for discourse at a given moment.

To clarify, a conceptualization (see Section 3.1.2.1.) at a given step is seen as dependent on, among others, neighboring usage events. In particular, the preceding usage event functions as the key reference point (Langacker 2001). To illustrate, while "[t]he default for cats is a domestic feline, the default for cat behavior is sleeping, and the default for mats is that they are spread out on the floor", "The cat is on the mat" does not have to evoke "a typical domestic feline reclining on a flat piece of woven material spread out on the floor" (Langacker 463–464). Instead, this usage event can refer to an unconscious animated tiger lying on a mat after a boxing match, if it follows, for example, a usage event concerning a cartoon.

Concurrently, Langacker indicates that usage events are influenced by cultural or social expectations. For instance, words considered profane in a church can be interpreted by interlocutors[95] as not appropriate in this context and thus they are not commonly used.

---

[94] As Nuttall (2018: 51) explains, discourse "may be seen to operate at the level of clauses, sentences, paragraphs, texts, etc", depending on "how closely we choose to examine it". However, "one significant level of structural organization (…) is that of attentional frames", or prosodic/punctuation units, according to the terminological solution adopted in this dissertation (see Section 3.1.2.2.3.).

[95] For Langacker, interlocutors can be "real or imagined". This assumption implies that writing can also be characterized with reference to the CDS. More specifically, when the production of, for instance, a research article progresses, each subsequent usage event is dependent on, among others, neighboring usage events.

Similarly, specific intonation in speech may be considered inappropriate in one context but essential in another, especially if a usage event is intended to serve as a warning. Hence, the proposal entails that the CDS is multilayered and involves "everything" (e.g. linguistic or paralinguistic expectations) that allows interlocutors to arrive at the same interpretation.

Given the usage-based nature of linguistic knowledge highlighted in C(C)G (see Section 3.1.), "any aspect of a usage event, or even a sequence of usage events in a discourse, is capable of emerging as a (…) unit" (Langacker 2001: 146). For example, the expression "Once upon a time" has gained a unit status and is thus widely recognized as the beginning of stories. Hence, when employed in a usage event set, this expression influences the interpretation of the other usage events as successive components of a given story. In other words, usage event sets (discourse constructs) elaborate or extend established, i.e. conventionalized, symbolic unit sets (discourse constructions) (see Section 3.1.2.2.1.).

The relevance of discourse conventions is also emphasized by Östman (2005), who demonstrates that "core" symbolic units (constructions) fail to sanction fully acceptable discourse usage events. More specifically, Östman (2005: 122) argues that the usage event "Mother drowned baby" is not "licensed by any of the 'core' constructions like the Determination construction, the Subject-Predicate construction, and the Phrasal Verb Phrase construction" which would instead collectively sanction a usage event "A mother drowned a baby", typical of, e.g., expository prose. However, "Mother drowned baby" is "clearly an acceptable construct [i.e. an acceptable usage event] in English" when functioning as a headline in a newspaper. Hence, Östman (2005) states that constructionist approaches should recognize the usefulness of discourse patterns[96], i.e. conventionalized form-meaning pairings roughly equivalent to discourse genres, as they

[96] The terminology used with reference to such complex constructions tends to vary (see Section 2.2.2.1.). For instance, Östman (2005) introduces the notions of, among others, "framing constructions" or "discourse constructions", claiming that they stand for pairings of genres and text types. On this account, with reference to, for instance, Swales (1990), Bhatia (1993), or Werlich (1976), Östman (2005: 131) emphasizes the distinction between genres and text types as genres are "contextual settings that are suggested on the basis of different activities that people engage in" (e.g. recipes or death notices) whereas text types are "defined on the basis of the manner in which sentences are organized as parts of a piece of discourse in relation to each other" (e.g. narrative, expository). Hoffmann and Bergs (2018), in turn, state that for cognitively-oriented research such a distinction holds rather minor significance. Hence, they apply the terms genres and text types interchangeably, together with the term roughly equivalent term "textual constructions". At the same time, Baicchi and Erviti (2018) apply the term "genre constructions". Therefore, for the sake of clarity, only the term "discourse genre constructions" is used with reference to such units in the present discussion, which roughly aligns with the proposal of Langacker (2008).

encapsulate certain information of what component structures are expected (and thus conventional) in a given genre.

4.1.2. Discourse genres

As Langacker claims (2008: 478), discourse occurs in various genres, e.g. job interviews, lectures, recipes, or linguistic articles, and individuals have "at least a rough idea of their typical properties". Although the notion of discourse genres has been theorized from various (complementary) perspectives[97], a more cognitively-oriented account stresses that a discourse genre construction[98] should be perceived as a set of schemas derived from encountered instances specifying, among others, what passages should be present in a given instance or how they are holistically structured.

As Langacker (2008: 480) notes, knowledge of a given genre constitutes a set of schemas reflecting "a recurring commonality in regard to some facet of their structure: their global organization, more local structural properties, typical content, specific expressions employed, matters of style and register, etc.". For instance, the pronouns "I and you are very frequent in conversation and personal letters, but they hardly occur at all in newspaper headlines" (Langacker 2008: 479). Likewise, Fillmore (1982: 117–122) states that knowledge of genres entails "knowledge about how to interpret particular passages in it, how to expect the text to develop, and how to know when it is finished. In the same vein, Baicchi and Erviti (2018: 576) state that a genre "can be defined as a holistic system (…) that is entrenched in the long-term memory of communicators". On this account, constructionist approaches aim to model discourse genres as "construction-like objects that are acquired in a usage-based fashion just like 'ordinary' constructions" (Hoffmann and Bergs 2018: 5), and can be characterized accordingly (see Section 3.1.2.2.).

In other words, apart from indicating that discourse genre constructions function, among others, as units that sanction discourse usage events, thereby relating such units to the fixedness-novelty continuum (see Section 3.1.2.2.1.), discourse genre constructions are also theorized in relation to the remaining continua, namely, the simplicity-

[97] See, for instance, Bhatia (1993) or Hyland (2007).

[98] Given the idea of the gradual emergence of constructions, the term discourse genre construction is applied with reference to emergent, emerging, and emerged discourse genre constructions. Hence, it is seen as roughly equivalent to the notion of speech event types introduced in Chapter 2, which is used by ELF scholars as an umbrella term for well-known (emerged) and less established (emerging) genres.

complexity continuum (see Section 3.1.2.2.3.) and the specificity-schematicity continuum (see Section 3.1.2.2.2.).

To begin with, although discourse genre constructions are complex constructions by definition, i.e. they may consist of words, sentences, paragraphs, or sections, the distinction between simple and complex constructions in relation to such units is still applied. As Bergs (2008) proposes, a headline can be seen as simple whereas a novel can be viewed as complex as it consists of several parts "which need to come together in order to result in the complex construction" (Bergs 2008: 274). In fact, as Bergs (2008) claims, many discourse genre constructions consist of sub-genres (sub-constructions), making the characterization of these constructions best represented in the form of hierarchical structuring. For instance, one of such discourse genre constructions can be a newspaper, which encompasses various sub-constructions, such as headlines or distinct sections, each contributing to the overall structure at a higher level of organization.

However, "[t]he higher the level of organization, the more likely it is for conventional units to be schematic" (Langacker 2008: 480). As Langacker explains, if we look at specific expressions learned as units, we notice that their quantity decreases as their size increases, i.e. "we learn thousands of words and phrases, many fewer clauses, still fewer complex sentences, hardly any lengthy passages, and probably not a single novel". Likewise, Hoffmann and Bergs (2018) argue that individuals often retain specific examples of songs or poems. However, when considering the Bible, its undeniably more complex nature reduces the likelihood of individuals memorizing it entirely.

Concurrently, whether a discourse genre is stored as specific or schematic largely depends on the input. In other words, if a person has continuous exposure to only one discourse genre construction, e.g. the recipe of how to cook pasta, this recipe will be stored as the specific pasta recipe construction. However, if a person has greater exposure to variants, it is likely that a more schematic recipe construction will develop. More precisely, from more specific constructions, such as the pasta recipe construction or the pancake recipe construction, the recipe construction emerges with a more schematic meaning of how to prepare a meal (Hoffmann and Bergs 2018).

Moreover, considering that any aspect of a sequence of usage events can become a unit, including information related to attentional frames, the complexity of discourse genre constructions is currently also theorized as a combination of inherently more specific (i.e. linguistic) and more schematic (i.e. paralinguistic) constructions (see Section 3.1.2.2.2.). The above tendencies in theorizing discourse genre constructions are

elaborated and extended through empirical studies in the next section to delineate the methodological approaches currently adopted and to reinforce the growing importance of the discourse level in C(C)G.

4.2. Researching discourse genres as constructions

As noted by Hart and Queralto (2021: 533), "any form of expression, whether visual, manual, or auditory, that features alongside language in a usage event has the potential to become part of a multimodal construction", i.e. a construction combining different modes. Broadly speaking, the notion of a mode in ~~a~~ cognitively-oriented research is defined as a "semiotic resource used in meaning-making such as images, writing, layout, music, gesture, speech, moving image, and so on" (Gibbons and Whiteley 2018: 249). Nevertheless, "what might count as a mode is an open-ended set" (Page 2009: 6).

For example, Forceville (2009) focuses on the sensory perceptual system to propose categories of modes, suggesting that modes can be associated with each of the five senses, e.g. the visual mode or the auditory mode, noting, at the same time, that such a classification may be too broad. Hence, Gibbons (2012), for instance, argues that the visual mode encompasses, among others, written language, images, and gestures and claims that finer-grained approaches should be applied to discern differences between them. In fact, since "multimodality as a field of academic research is still at an embryonic stage" (Gibbons 2012: 8) and the categorization of modes is rather unclear, it is suggested that research should focus on the establishment of which modes are preferred in specific contexts (Page 2009).

Therefore, to demonstrate how paralinguistic constructions are incorporated into research on discourse genre constructions, the ensuing discussion focuses on established "preferences" identified by researchers regarding the co-occurrence of modes. In other words, as a starting point for further considerations, Langacker's (2008) proposal that discourse genre constructions can be divided into spoken-language genres and written-language genres is adopted. Also included are observations by Gibbons and Whiteley (2018) as well as by Hart and Queralto (2021), who point out that "when we talk, we rely on the modes of spoken language, intonation, and gesture" (Gibbons and Whiteley 2018: 249) and when we write, we rely on the modes of written language, punctuation and images (Gibbons and Whiteley 2018; Hart and Queralto 2021)[99], among others.

[99] Viewed in this way, modes can be regarded as roughly equivalent to Langacker's (2001, 2008) channels (see Section 3.1.2.2.2.).

Therefore, the next two sections present discourse genre constructions (henceforth also DGxCs) in spoken (Section 4.2.1.) and written (Section 4.2.2.) modes, i.e. the "core" modes (Langacker 2008), highlighting their (potential) interplay with co-occurring modes.

4.2.1. DGxCs in the spoken mode

In order to indicate both prevailing and evolving directions in research on DGxCs in the spoken mode as well as to align the current section with the overall architecture of Chapter 2, studies presented in this section are divided into those concerning non-academic and academic DGxCs. The discussion centers on providing an overview of research on both "specific passages" in DGxCs and "their more global organization" (Langacker 2008), highlighting an interplay between core and non-core modes, where relevant.

4.2.1.1. Non-academic DGxCs

To begin with, one of the analyses concentrating on non-academic DGxCs is the study conducted by Langlotz (2015). In his proposal, Langlotz aims to develop a model accounting for both the emergent nature of meaning construction during conversations and the stable expectations arising from a language-user's knowledge of a given DGxCs. To achieve this aim, Langlotz integrates Langacker's theoretical assumptions regarding the CDS (see Section 4.1.1.) with Barsalou's model of simulators and simulations (see Section 3.1.1.3.), thereby proposing genre-simulators (frames) encompassing, among others, knowledge about specific lexical choices or the sequential organization of a given DGxC and thus guiding relevant genre-based usage-events.

The analysis illustrating Langlotz's proposal is based on extracts representing tourist-information transactions, which are seen as "a speech genre with stable discursive norms that impose constraints on allowable speech turns, lexical choice, syntax, and speech style" (Langlotz 2015: 517). For instance, one of the extracts concerns the management of the booking-a-hotel-room activity. Firstly, the focus is on the expected sequence underlying the DGxC, i.e. an entrenched/conventionalized attribute-value set within a frame, depicted as "a nested frame-structure" (see Figure 11). The nested structure reflects the fact that as the conversation progresses, the employee, for instance, gradually activates additional relevant concepts necessary for communication. In other words, at the beginning of the interaction, the employee expects to be responsible only for providing service in general but then he/she focuses on more specific knowledge

related to booking a room and gradually shifts to more detailed concepts (i.e. values of the SERVICE attribute).

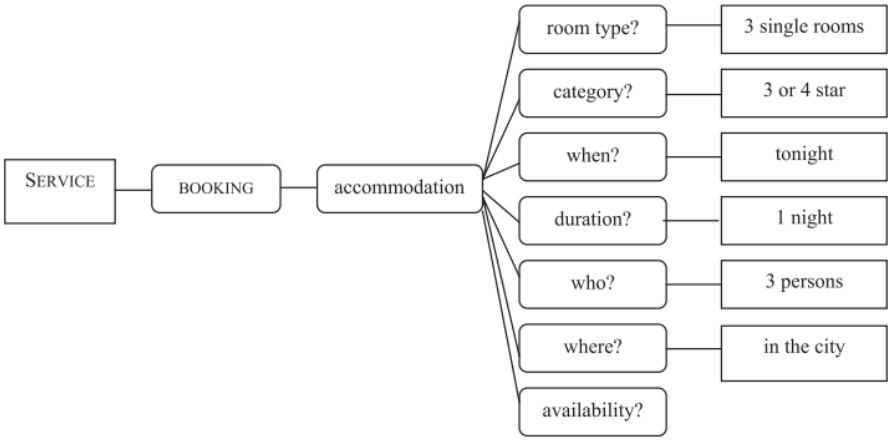


Figure 11. Situated conceptualization (Langlotz 2015: 533)

As Langlotz (2015: 534) puts it, this form of representation shows "the construction of their joint conceptualization process by concretizing the hotel-room category step-by-step".

Secondly, the focus is on the departure from the expectations underlying the BOOKING frame, which requires recognizing that the conversation takes place when there is an important event in the city; hence, hotels are fully booked. Langlotz (2015) observes that when the employee uses the word "outside", which is not associated with the WHERE value of the BOOKING attribute, it leads to a misunderstanding between the hotel employee and the guest (see Figure 12).

What is the cheapest room tonight?
Do you have much accommodation?
There is no room available
Maybe you get an accommodation at the YMCA
Or a youth hostel?
Or the youth hostel but I'm afraid it's booked
the cheapest one is somewhere outside
Outside as in … outside the city you mean?
No [LAUGHS] outside of the rooms … outside … there is no hotel available

Figure 12. Conversation (adapted from Langlotz 2015)

As Langlotz (2015: 540) explains, the employee "uses 'outside' as a cue to activate a different concept-simulator in order to evoke a simulation [a conceptualization] of the tourist sleeping in the street". However, the tourist appears unable to imagine the scenario

of sleeping there because the conventional genre-simulator concerning booking a room limits his/her interpretative possibilities at this moment (see Figure 13).

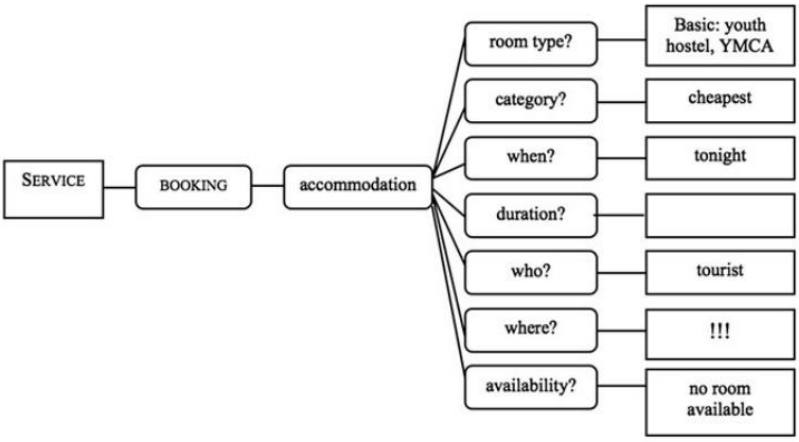


Figure 13. WHERE value

At the same time, Langlotz acknowledges that also facial expressions or gestures play a significant role in DGxCs but does not elaborate on these non-core modes in his model of meaning construction.

Unlike Langlotz (2015), Hirrel (2018) incorporates non-core modes in her analysis. More specifically, the author investigates the relationship between speech and cyclic gestures, i.e. "manual co-speech gestures that are characterized by a circular movement of the hand or arm" (Hirrel 2018: 9), in the talk show genre. In total, 501 instances of cyclic gestures are collected for this investigation, which are derived from fragments of 95 distinct American English talk show episodes, comprising approximately 12 hours of video.

The study is conducted in several stages. Firstly, Hirrel (2018) develops separate classifications to describe speech and gestures. As the author indicates, this stage requires several iterations to introduce necessary refinements resulting from continuous analysis. Regarding gesture classification, Hirrel (2018) decides to describe their forms based on multiple variables, such as handedness (unimanual, bimanual) or hand distance (near, wide, neutral), the selection of which stems from the fact that they are recognized by other scholars as significant for research on sign language and gestures. Subsequently, the researcher develops a classification for linguistic constructions, limiting the description to those that co-occur with gestural constructions. The classification involves dividing them into more or less complex units, e.g. words, phrases, and clauses, and identifying their functions.

After developing the necessary classifications, Hirrell (2018) firstly concentrates on a more quantitatively-oriented approach. In other words, the researcher focuses on the establishment of the frequency of co-occurrence of gestural and linguistic constructions and admits that a more qualitatively-oriented approach is limited to the constructions that, among others, can be seen as frequently co-occurring (e.g. at least 5% of the total) (see Section 6.1. for details). Considering the established variables for describing the constructions under study, Hirrel (2018) investigates whether there is any significant relationship between the categories. In other words, the author explores whether there is a significant association between, for instance, the variable of handedness and its occurrence with pronouns. To achieve this, Hirrel (2018) relies on several tests to determine statistically significant differences, including the chi-square test and, when the chi-square test cannot be applied, Fisher's exact test[100].

Secondly, the author adopts a more qualitative approach and aims to propose how such a complex (or multimodal) usage event, i.e. an assembly elaborating and extending linguistic and gestural constructions, can be represented. To illustrate her proposal, the researcher describes one gestural form to which the meaning "to present something as topical for an interactional purpose" (Hirrel 2018: 183) is assigned and combines it with the form and function of co-occurring linguistic expressions (see Figure 14; relevant linguistic expressions are bolded).

Figure 14. Linguistic and gestural expressions (adapted from Hirrel 2018)

Hirrel (2018: 183) claims that "He's gonna be the nominee" is used to establish a "specific topic that the speaker wants to offer for joint activity". At the same time, "I mean did you ever think you would get to this point", elaborates the semantic pole of the preceding

[100] The discussion concerning these tests is provided in Section 6.1. as the study described in the present dissertation follows the same approach.

construction. In other words, it serves as a question; hence, in relation to the previous statement, it indicates that the speaker wants the listener to engage with the topic introduced. A similar function, i.e. "to present something as topical for an interactional purpose" is attributed to the gestural form.

Importantly, following Langacker's (2001, 2008, 2017b) proposal (see Section 3.1.2.2.2.), Hirrel (2018) represents the phonological and semantic poles of such a complex (or multimodal) symbolic assembly (see Figure 15 below) as a combination of linguistic (vocalization) and paralinguistic (gesture, intination) structures.

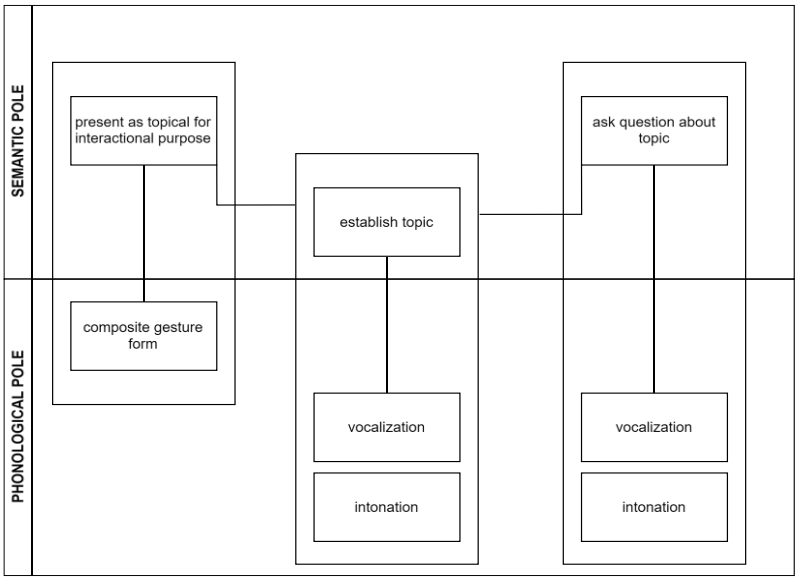


Figure 15. Multimodal symbolic assembly (adapted from Hirrel 2018: 183)

Simultaneously, Hirrel (2018) admits that her analysis is still rather coarse-grained in that it does not, for instance, fully account for the role of intonation, and emphasizes the need for further research on modeling relations between modes.

4.2.1.2. Academic DGxCs

When it comes to academic DGxCs, a good illustration of the C(C)G framework is provided by Baicchi and Erviti (2018), who aim to show that a given DGxC can be associated with a specific set of constructions. More precisely, the analysis concerns academic lectures derived from the MICASE corpus (see Section 2.2.2.2.) and intends to delineate constructions used to express complementary contrastive meaning.

As Baicchi and Erviti (2018: 581) explain, in contrastive constructions, e.g. "X or Y", "two segments of discourse simply oppose each other and cannot coexist". In

complementary contrastive constructions, in turn, e.g. "X although Y", "even though the elements compared are contradictory, the existence of the first element does not preclude the existence of the second, and vice versa". In essence, the analysis involves developing a comprehensive list of such constructions and examining which of them appear in academic lectures.

To establish such a list, the authors rely on explicit norms (see Section 3.1.2.2.1.). In other words, they firstly conduct a thorough investigation of related studies within a broadly-understood Construction Grammar approach[101]. Secondly, they search for synonyms in dictionaries to identify as many constructions with related meanings as possible. Moreover, recognizing that certain forms encoding complementary contrastive meaning are not listed in dictionaries, the researchers additionally decide to rely on their knowledge of native languages (i.e. Basque, Italian, and Spanish) to identify other potential candidates.

Once they gather a set of constructions, the authors compare their definitions and those provided by dictionaries and discover that many dictionary definitions treat such constructions as interchangeable, which is not supported by their preliminary study. To clarify, the authors observe that, in some contexts, one construction is more frequently used, while others are rare or have slightly different meanings. Hence, following Langacker's (2013) proposal concerning active zones[102], they decide to establish seven sub-types of the construction under study, i.e. neutral, concessive, correcting, topic-changing, topic-avoiding, refusal-apology, and evaluative; however, there are only four types which appear in academic lectures, i.e. neutral, concessive, correcting, and topic-changing. This observation is explained by the fact that the sub-types under study do not comply with the semantic pole of the academic lecture. For instance, as the authors explain, evaluative constructions inherently involve a level of uncertainty or hesitation, conflicting with the meaning of confidence and reassurance that expert speakers are expected to convey during academic lectures. In this way, as the authors conclude, it is shown "how particular genres promote the use of certain constructions" (Baicchi and Erviti 2018: 597).

[101] In other words, the authors do not specify whether they rely on Cognitive (Construction) Grammar, Cognitive Construction Grammar, etc.

[102] The notion of active zones is equivalent to the idea of activating relevant parts of the frame to understand a given situation. For instance, as Langacker (2013) explains, in the context of biting, the active zone of a dog consists primarily of its teeth.

4.2.2. DGxCs in the written mode

In accordance with the division adopted in Section 4.2.1., DGxCs in the written mode are presented with reference to non-academic and academic DGxCs. Likewise, this discussion focuses on presenting research that either examines specific constructions within DGxCs or presents DGxCs in a more holistic manner, i.e. in relation to typical sequences of co-occurring constructions. Analogically, the section illustrates both studies pertaining to constructions in the written mode alone and those accounting for their co-occurrence with non-core modes. At the same time, due to the recognized need for further research on written academic ELF (see Section 2.3.), to be addressed in Chapters 5 and 6, the discussion of DGxCs in the written mode is presented in a more detailed manner than considerations realated to the spoken mode (see Section 4.2.1.).

4.2.2.1. Non-academic DGxCs

To begin with, one of the studies concerning non-academic DGxCs is the exploration by Hoffmann and Bergs (2018), who concentrate, among others, on describing Shakespearean sonnets as DGxCs. In their approach, the authors admit that the description of DGxCs concerning their graphological poles may include not only "individual linguistic elements (morphemes, words, phrases, paragraphs) but also contextual features, such as book cover, shape or size, layout, their place at the bookstore, etc." (Hoffmann and Bergs 2018: 12). Likewise, the description of their semantic poles may involve the presentation of not only "frames and frame elements" but also, among others, connotational meaning, e.g. aesthetic or emotional attributes, as well as social and contextual constraints.

Hoffmann and Bergs (2018) base their analysis on the idea that the diversity of Shakespeare's sonnets that people are exposed to influences the schematic construction arising based on this input. As they note, some sonnets may be stored in the long-term memory as specific constructions resulting from continuous exposure to the same sonnet. At the same time, they acknowledge that if exposure to a given sonnet is not frequent enough, it is rather unlikely for a user to store the entire sonnet in his/her long-term memory. Instead, in such cases, information about sonnets' graphological and semantic poles is probably limited to their most salient features, such as information about rhymes.

At the same time, they propose that if a given user is exposed to numerous Shakespearian sonnets, it is likely that they store a more schematic construction in the long-term memory, which is formed by identifying features that "consistently re-occur"

in all the sonnets he or she is exposed to (Hoffmann and Bergs 2018: 10). As a consequence, such a schematic construction, can be characterized rather broadly, i.e. "descriptions such as 'has fourteen lines', 'the first twelve rhyme ABAB, CDCD, EFEF' or 'the last two rhyme' will often come up. This abstract FORM template of the Shakespearean sonnet construction will, for lay readers, be linked to an abstract MEANING pole that identifies the function of this (…) construction as expressive of one's love" (Hoffmann and Bergs 2018: 11). Additionally, this construction will also be linked to various social and contextual aspects, including a certain formality and solemnity, as well as the types of situations in which people might consider it appropriate to recite the sonnet.

While Hoffmann and Bergs (2018) only declare the importance of combining core and non-core, e.g. layout, modes in DGxCs but fail to present specific solutions, Östman (1999, 2005) succeeds in both respects. To begin with, the researcher notes that a description of a DGxC may involve not only "conventionalizations of specific linguistic properties" but also, or in particular, "the visual, graphic display" of a DGxC (Östman 2005: 121–133). As Östman (1999) illustrates, a recipe, for instance, involves providing information regarding the ingredients in a list-like manner, followed by explanations of the sequential steps of how to prepare a meal. At the same time, Östman (1999: 81–83) highlights that "[t]he most noticeable feature of recipes is, concretely, the graphical layout they conform to". In other words, when someone mentions, for example, a recipe, "the first thing that comes to mind (in our culture), in terms of which we most readily, most immediately, and most efficiently understand the notion of a recipe is the visual, graphic display" (see Figure 16).

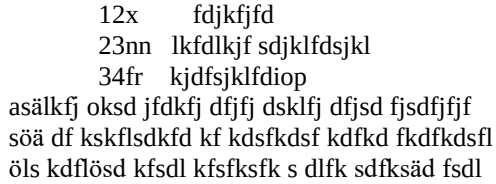


Figure 16. Recipe image (Östman 1999: 82)

Moreover, Östman (1999: 90) emphasizes that when cooking instructions have alternative "physical shapes", it necessitates additional cognitive effort to understand and classify such non-typical cooking instructions as recipes. On this account, it may be said

that a typical shape associated with recipes (see Figure 17)[103] is "not only linked to perception, but to cognition" as this form activates relevant frames (see Section 3.1.1.3.2.) for comprehension.

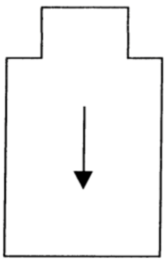


Figure 17. Recipe pattern (Östman 1999: 83)

At the same time, Östman (1999) claims that such expectations associated with a given DGxC do not have to be universal (see Section 3.1.2.2.1.). In other words, while recipes have a pattern which is common for many communities, the development of a universal one for, e.g., death notices may be challenging. As Östman (1999: 84) illustrates, "whereas in Britain, deaths are recorded as text, in Finland deaths are advertised much more graphically". To be more specific, British death notices can simply be characterized by a typical sequence of providing information (see Figure 18).

Died a few days ago, at his house in Greenwich, Capt.
Robert Walter, of the Royal Navy.

Figure 18. British death notice (Östman 1999: 84)

Finish death notices, in turn, are characterized by a cross (or another religious symbol) positioned to the left, which is followed by the dead person's name in bold. This information is also accompanied by the dates and places of birth and death, names of the mourners, often a psalm, and lastly, details regarding the location of the funeral (see Figure 19).

[103] This representation is, in fact, similar to the depiction of maximal goodness pertaining to symmetry (see Figure 5).

Rakkaamme, insinööri

Pasi Antero
Ailio

s. 9. 7. 1951 Kajaanissa
k. 7. 10. 1996 Naantalissa vaikean sairauden murtamana

Kaipaamaan jäivät
Marja-Helena, Heini ja Henri
Lea-äiti
Muut sukulaiset ja lukuisat ystävät

Sinun suunnaton rakkautesi ja huolenpitosi.
Rakas Pasi auttaa minua jaksamaan ja jatkamaan meidän yhdessä aloitettua purjehdustamme nyt elämän vesillä.
Yhteisissä lapsissamme Heinissä ja Henrissä olet aina oleva rinnallani.
Odota minua siellä jossain.
Kiitän yhteisistä ihanista vuosista Rakas
Marrusi

Jumalalle kiitos poikani elämästä!
Äiti

Siunaustilaisuus Naantalin kirkossa lauantaina 26.10.1996 klo 10.00. Sen jälkeen muistotilaisuus Naantalin Merisalissa, jonne lämpimästi tervetuloa.

Figure 19. Finnish death notice (Östman 1999: 85)

As a consequence, Östman (1999) states that a Finn might not immediately recognize or even consider a British death notice as a proper death notice since such an instance is, in fact, an example of extension which does not align with the expectations established within the Finnish community.

Another approach highlighting an interplay between core and non-core modes is developed by Hart and Queralto (2021), who aim to show that the two modes can evoke the same construal of a given situation. As the authors (2021: 529) put it, they aim to indicate "that images and language usages which are proximal to one another in a multimodal text can be expected to exhibit the same or consistent construals of the target scene". Building on Langacker's approach (see Section 3.1.2.1.), Hart and Queralto (2021) represent the construction as a multimodal symbolic unit, emphasizing, at the same time, that the graphological pole consists of two parts, e.g. a word and an image, which converge in one semantic pole (see Figure 20).

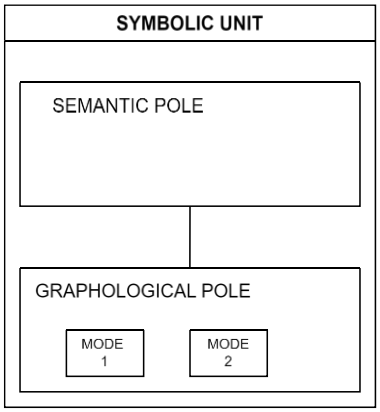


Figure 20. Multimodal symbolic unit (adapted from Hart and Queralto 2021)

To illustrate examples from which conventionalized multimodal symbolic units can emerge[104], Hart and Queralto (2021) focus on analyzing news photographs and their captions. As a common denominator to describe the semantic pole in the two modes, image schemas (i.e. basic concepts) are utilized. The authors rely on, among others, Talmy's proposal closely related to Gestalt schematic structures (see Section 3.1.1.2.). More precisely, Hart and Queralto (2021) focus on describing plexity, i.e., "a quantity's state of articulation into equivalent elements" (Talmy 2000: 48) with values uniplex (one) and multiplex (more than one), as well as boundedness, i.e. possessing a limit (see Section 3.1.2.2.2.), with values bounded (with a limit) and unbounded (without a limit). For instance, the authors indicate that plexity and boundedness can be employed with reference to the following image and its caption (see Figure 21).

The huge column of migrants passes through fields in Rigonce, Slovenia, after having been held at the Croatia border for several days.
MailOnline, 25 October 2015

Figure 21. Plexity and boundedness

As the authors indicate, the phrase "column of migrants" portrays the group as a single, oblong shape (and thus uniplex). The image, in turn, indicates uniplexity by presenting a high degree of closeness among migrants, forming a single oblong shape corresponding to the linguistic description. When it comes to unboundedness, the authors point out that the use of the adjective "huge" suggests that the column is of significant magnitude, making it perceivable as having no visible end. In the image, the unbounded value is depicted, as the limit of the column is not included in the picture.

[104] The authors claim that the exploration of (emerging) multimodal constructions necessitates a quantitative approach. In other words, it is crucial to determine how frequently the two modes co-occur, suggesting a multimodal construction.

Another study that explores an interplay between modes is the analysis conducted by Zenner and Geeraerts (2018), who seek to establish "the cornerstones and building blocks" of the Image Macro Construction, i.e. a DGxC which contains a sequence of words placed in the upper and bottom side of the image. The data are collected by means of a number of platforms, such as KnowYourMeme.com, Reddit.com, or Google Images.

Based on the analyzed examples, the authors aim to propose a schematic representation of the Image Macro construction (see Figure 22). As Zenner and Geeraerts (2018: 175) explain, "[t]he grey boxes in the figure represent directly visible parts of the meme. The black box is used for an implicit element, not visible in the image macro. White is used for analytic categories".

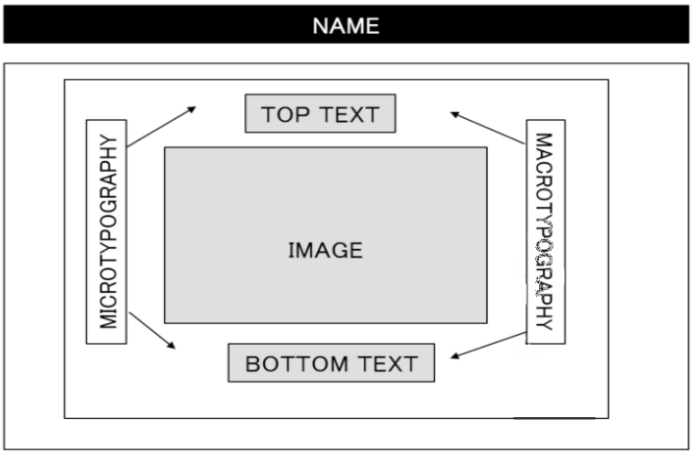


Figure 22. Image Macro Construction

In a more detailed discussion, the authors indicate that the first vital component of the construction is the image. However, while it seems obvious that an image macro necessitates an image, the specific kind of image utilized in image macros displays diversity. Although it is not indicated in the representation, their semantic pole is presented as modifying the meaning of constructions pertaining to the written mode. In other words, the authors indicate that a given image determines which frame should be activated for the correct interpretation. To clarify, the proper interpretation of "One does not simply walk into Mordor" (see Figure 23) seems to be possible only when one considers the fact that the Image Macro construction is created based on the specific image and the line taken from the film adaptation of *The Lord of the Rings*. In this scene, the character Boromir talks about the extremely difficult task of walking into Mordor,

saying “one does not simply walk into Mordor”. The meme uses this line to suggest that, in this case, resisting bacon should be seen as similarly impossible and destined to fail.

Figure 23. Meme

The second vital component is presented with reference to linguistic constructions which are, regarding macrotypography[105], divided into top and bottom texts. As the authors explain, while the top text establishes the context, the bottom text serves a role akin to the punchline in traditional jokes. At the same time, they emphasize that one of the most consistent features of this construction pertains to microtypography, i.e. top and bottom texts are commonly displayed in the “Impact” font[106], which is typically utilized in uppercase letters, featuring white text with a black contour. Nevertheless, although the authors emphasize that layout and types of font are an integral part of the form pole of the construction, they do not indicate whether (and if so how) such forms contribute to the meaning of the construction, i.e. whether and how they can be paired with the semantic pole

Both layout and types of font are considered in the analysis conducted by Fisher and Aarestrup (2021), who present Instagram posts as instantitions of a DGxC. More specifically, following Bateman et al. (2007), Fisher and Aarestrup (2021: 3) claim that when Instagram posts are analyzed it should be taken into account that “layout features can serve as genre markers and indicate the kind of text produced”. Hence, in their view, the Instagram post functions as a discourse genre construction as it has a unique spatial arrangement and specific parts, which, taken together, make the instances easily recognizable as the Instagram post construction. More specifically, each post is said to be

[105] Macrotypography is equivalent to macropunctuation (Waller 1980) (for a definition of punctuation see 3.1.2.2.2.). In the case of image macros, macropunctuation refers to layout.
[106] As the authors indicate, Birdeau and Berret (2014) claim that this font is now treated as the “meme font”.

made up of an image, the company's logo and name, a caption containing text and hashtags, a section of comments, a set of action buttons (allowing users to like, share, or comment on the post), optional geo-location, and the total number of likes (see Figure 24). On this account, the unique layout of Instagram posts includes both (optional and obligatory) content elements and interface elements, making the posts "a tapestry of many voices" (see Section 2.1.1.). In other words, this DGxC emerges as the combination of the work of the platform's developers and the content provided by users, who modify the immutable sequence of components by adding personal data such as photos or captions..

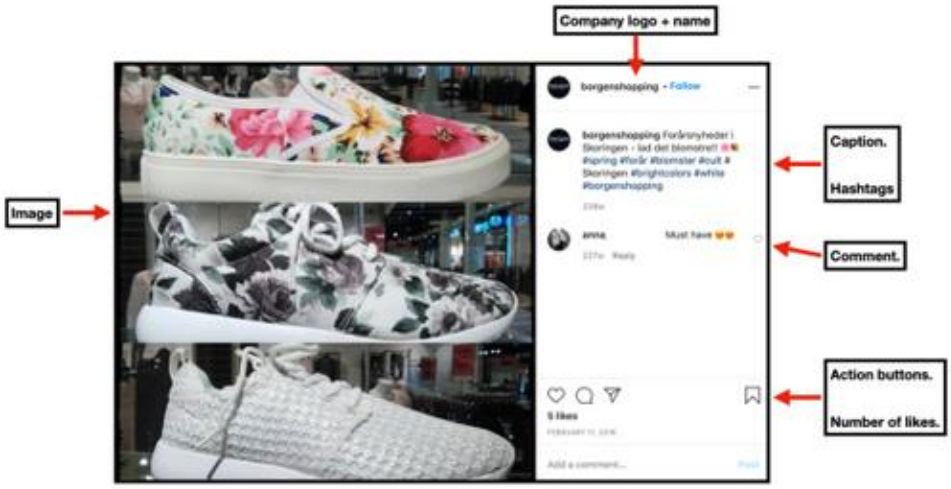


Figure 24. Instagram post

As Fisher and Aarestrup (2021: 1) state, "[w]e find that Instagram posts may constitute stable configurations (…) Nevertheless, considerable variation in the realization of Instagram posts can be observed". Therefore, in their analysis, they focus on evaluating the repetitiveness of features in the posts.

The data used for the analysis represent 70 Instagram posts by Danish and American department stores. To select Danish department stores, the authors identified the Instagram profiles of the largest department stores in each of Denmark's five regions and selected the top three stores with the highest follower counts for analysis. Furthermore, three American department stores were selected based on the list by TOTEMS (2014), which ranks the most popular department stores according to their follower count and the number of posts under their specific hashtags.

By modifying a model proposed for describing websites (Seizov 2014), they develop a coding scheme tailored to the Instagram platform to describe images and

captions separately. In other words, they develop a set of attributes and their values to evaluate selected parts of the Instagram posts. For instance, they determine the main motif (with such values as object or person), topic (with such values as advert or celebrity), or speech acts (with such values as representatives or directives). To evaluate the repetitiveness of their values, they establish three possible groups: flexible (no value is preferred), restricted (two or three values are preferred), and fixed (one value is preferred and no or almost no other values occur).

The findings reveal a general trend, i.e. the US department stores exhibit more restricted choices compared to the Danish department stores, which shows that the Instagram post construction can be community-specific. Moreover, based on the analyzed examples, Fisher and Aarestrup (2021: 20) aim to establish "an abstraction over all Instagram posts" which is "akin to postulating a general constructional schema that allows people to produce 'well-formed' [i.e. conventional] posts" (see Figure 25).

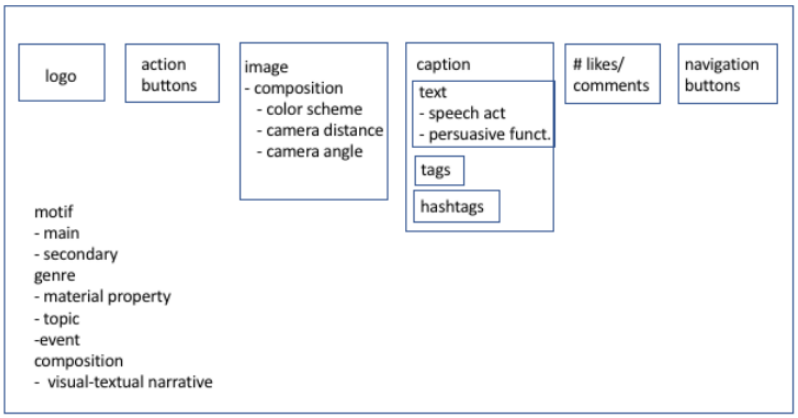


Figure 25. Instagram post construction

Nevertheless, as Figure 25 shows, this representation does not clearly indicate the semantic and graphological pole of the construction under study.

4.2.2.2. Academic DGxCs

When it comes to academic DGxCs, one of the preliminary studies is conducted by Wiedemer and Vieira (2022), who aim to present the academic abstract as a DGxC. According to Wiedemer and Vieira (2022: 272–273), the abstract, also referred to as the academic-scientific summary, is a condensed representation of the research, which should be seen as "a form-function pairing stored in the mind of those who have the experience

of reading or writing texts at the university", e.g. research articles, dissertations, or conference proposals.

Their analysis concentrates on exploring "specific passages" (see Section 4.1.2.) that can be associated with this construction. Hence, the authors indicate that the abstract is associated with the first or the third-person construction, while the second-person construction is less common. Concurrently, they indicate that constructions with verbs in the present or past tense are the most common. More specifically, such constructions are typically found in the initial section of the abstract, which focuses more on the main ideas of the text the abstract is supposed to encapsulate. However, these constructions can also appear when presentingthe results and conclusions, and occasionally when describing the methodology. Additionally, the authors claim that it is quite common to use hedges.

Simultaneously, Strugielska and Guttfeld (2024) present an alternative approach to analyzing the abstract as a DGxC, suggesting that it can be examined in its more elaborated form. Building upon the assumptions of C(C)G, the authors propose that the abstract, as a construction, is capable of mixing with other DGxCs in its close vicinity on the website or in the PDF format. In other words, they argue that the digital environment in which abstracts of research articles appear facilitates blurring the boundary between the abstract and its context, leading to clusters of constructions which may constitute a higher-level assembly. In essence, they advocate identifying spatial groupings (see Section 2.2.2.2.1.) composed of abstracts and other DGxCs, utilizing paralinguistic cues to discern potentially more complex, or elaborated (see Section 3.1.2.2.1.), constructions. Thus, they divide abstracts into two types, i.e. basic (a DGxC conventionalized within the academic domain) and elaborated (a new type of abstract consisting of the basic abstract and co-occurring constructions).

Drawing upon a dataset comprising 2136 basic and 2108 elaborated abstracts sourced from the DISCOWER corpus (see Chapter 5), the authors undertake a detailed examination of hyperlinks within them across three disciplines. These hyperlinks are categorized into those facilitating interconnectedness (URL links) and those promoting interactivity (email links). The findings of this analysis reveal disciplinary distinctions in the utilization of such links. Consequently, the study underscores that the emergence of knowledge pertaining to abstracts, whether in their basic or elaborated forms, may be contingent upon discipline-specific experiences.

4.3. Concluding remarks

To sum up, Chapter 4 sought to introduce the C(C)G approach to discourse genres. The first section emphasized the significance of incorporating discourse genres into constructionist accounts while the second one outlined existing methodological approaches to studying DGxCs, focusing primarily on DGxCs in the written mode.

The discussion indicated that although DGxCs are characterizable with reference to the continua of features introduced in Chapter 3, i.e. fixedness-novelty, schematic-specificity, and simplicity-complexity, such characterization is mainly achieved with reference to selected aspects. In other words, the above overview suggested that current studies concentrate on, for example, presenting existing convention, delineating new DGxCs, or establishing more schematic constructions based on more specific ones. Concurrently, such studies focus on describing DGxCs with reference to constructions pertaining to one mode or multiple modes (and hence presenting DGxCs as more or less complex) as well as indicating that constructions pertaining to one mode can be combined, forming higher-level assemblies.

However, the discussion also indicated that certain areas seem to require further research and clarification. Firstly, while non-academic DGxCs are explored more holistically, considering typical sequences of constructions that constitute a given DGxC and their more multimodal nature, academic DGxCs are still predominantly studied with respect to specific passages and constructions in the (core) written mode. Secondly, while aiming to offer a way of representing DGxCs, researchers provide a rather coarse-grained perspective on DGxCs as they tend to overlook the potential impact of social and cultural experiences that can lead to the emergence of community-specific discourse genre constructions.

At the same time, the suggested representations of DGxCs appear rather heterogeneous as researchers focus on either the semantic pole or the graphological pole, or offer (limited) integrated perspectives that underscore selected relationships between modes. Thus, for instance, the status of paralinguistic constructions, such as prosody and punctuation, in DGxCs remains underexplored, with researchers often acknowledging their presence but rarely incorporating them as components of form and/or meaning poles of DGxCs.

Hence, it appears that further advancements concerning DGxC within a C(C)G approach can be achieved when more holistic (and multimodal) approaches to academic

DGxCs in the written mode are offered, and when the role of paralinguistic constructions is clarified and/or taken into account.

This direction of research is, in fact, possible by utilizing the DISCOWER corpus (see Section 4.2.2.2), the compilation of which involved developing a method for identifying a potentially new DGxC attributable to academic ELF communities. Therefore, the next chapter focuses on a detailed description of this corpus as it functions as the source of data analyzed in Chapter 6.

Chapter 5

The DISCOWER corpus: towards the elaborated abstract construction

5.0. Outline

The purpose of Chapter 5 is to present the principles of compiling the DISCOWER corpus, i.e. the source of data analyzed in Chapter 6. The chapter concentrates on explaining how the assumptions of Linear Unit Grammar (see Section 2.2.1.1.) and Construction Grammar (see Section 2.2.1.2.) employed in theorizing and researching academic ELF are integrated through and with the cognitive (see Section 3.1.1.) and constructionist (see Section 3.1.2.) commitments of Cognitive (Construction) Grammar to facilitate the discovery of an emerging academic discourse genre construction, i.e. the elaborated abstract construction.

In accordance with the general principles of corpus compilation discussed in Chapter 2, the creation of DISCOWER is described in Section 5.1. with reference to four aspects, i.e. corpus units, external criteria, corpus balance, and methods of data description. To be more specific, Section 5.1.1. provides a detailed explanation of the assumptions shaping the definition of corpus units and indicates how the tokens were identified. The second subsection discusses factors facilitating both the recognition of tokens attributable to academic ELF disciplinary communities and their consistent description (see Section 2.1). Additionally, the section presents the criteria for choosing academic ELF disciplinary communities from which the tokens were selected. Next, the third subsection presents quantitative data concerning the DISCOWER corpus and offers an interpretation of its balance. The final subsection, in turn, details how the data were described, emphasizing linguistic and paralinguistic constructions by means of which the tokens can be explored.

The chapter concludes with a summary of the key aspects guiding the process of corpus compilation and highlights the importance of further explorations of the potential elaborated abstract construction to establish a new DGxC among academic ELF disciplinary communities.

5.1. The DISCOWER corpus: compilation principles

The DISCOWER corpus, i.e. the discipline-oriented construction-based corpus of written English as a lingua franca, is the outcome of the DISCOWER project, in which I have been involved since 2021. The corpus contains data representing written abstracts of research articles in the PDF files, which were published from 2018 to 2021 by the members of three academic disciplinary communities, i.e. law, linguistics, and literary studies. The rationale behind creating the DISCOWER corpus stemmed from current usage-based research on academic ELF (see Section 2.2.2.) and construction-based approaches to (academic) discourse genre constructions (see Section 4.2.), the analysis of which led us, i.e. the DISCOWER project team[107], to the idea of providing a written ELF database that could be a valuable resource for exploring both of the aforementioned areas.

To ensure consistency among corpus compilers and minimize the overall disposability of the corpus (see Section 2.2.2.1.), the DISCOWER project team's efforts centered on establishing clear guidelines for the process of corpus compilation. Drawing inspiration from the general principles of corpus creation followed by academic ELF corpora (see Section 2.2.2.1.) and introducing solutions resulting from the theoretical background the project relied on (see Sections 3.1.1. and 3.1.2.), we delineated the principles including defining corpus units, discussing external criteria, outlining corpus balance as well as presenting the methodology used for data description. In the following subsections, a detailed presentation of our specific assumptions and the ways in which they were realized is provided, thereby elucidating the guidelines underlying the creation of the DISCOWER corpus.

5.1.1. Corpus units

To begin with, addressing the need for the creation of written ELF corpora as well as acknowledging the pivotal role academic disciplines play in academic ELF research (see Section 2.2.2.), the DISCOWER corpus includes written ELF abstracts of research articles. More specifically, the choice of this corpus unit resulted from our intention to provide a database complementary to WrELFA by incorporating a (previously unavailable) discourse genre construction, i.e. the abstract of research articles, closely tied to the one already available in WrELFA, i.e. research article[108].

[107] The list of team members is available at: https://discower.umk.pl/pages/main_page/.

[108] However, it should be noted that the authors of WrELFA do not define their corpus units as constructions in the Cognitive (Construction) Grammar sense (see Section 2.2.2.1.2.).

Secondly, recognizing the overall necessity to develop a constructionist approach to written academic ELF beyond its infancy (see Section 2.3.), and an ensuing particular need to employ usage-based approaches beyond Linear Unit Grammar (see Section 2.2.1.1.) and Construction Grammar (see Section 2.2.1.2.), we created a corpus grounded in the so-far underexplored in ELF framework of Cognitive (Construction) Grammar.

To be more specific, we perceived abstracts as form-meaning pairings, i.e. symbolic assemblies, arising from use by means of cognitive processes, e.g. grouping and reification, and further describable as, among others, more or less simple. In fact, the simplicity-complexity continuum (see Section 3.1.2.2.3.), with its focus on the relationship between the base and its elaboration, led to proposing basic and (potential) elaborated abstract constructions. To clarify, in accordance with the idea that elaboration depends on and provides details about its base, the (potential) elaborated abstract construction (henceforth EA) was perceived as necessarily including the basic abstract construction (henceforth BAxC) but not the other way round (see Figure 26).

Figure 26. Basic and elaborated abstracts

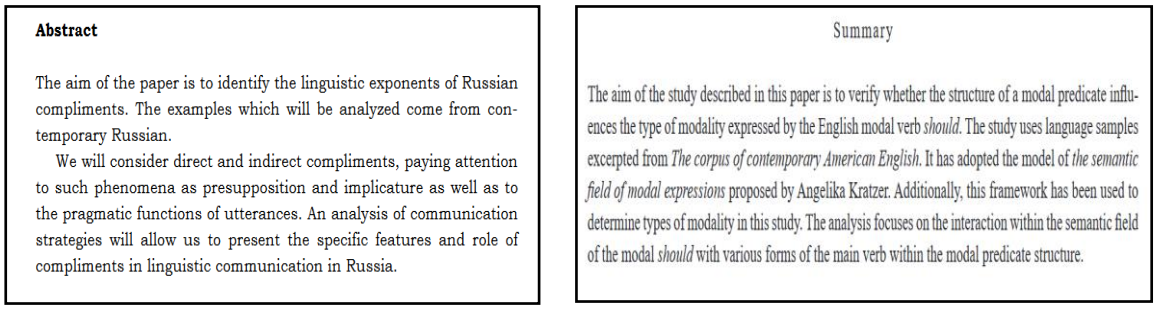
Importantly, the two abstracts were seen as occupying different positions along the fixedness-novelty continuum (see Section 3.1.2.2.1.). More specifically, when it comes to the BAxC, i.e. a paragraph or a set of paragraphs encapsulating the content of an academic research article (see Section 4.2.2.2.), it was seen as well-known among (and

regulated by the norms of) academic disciplinary communities. Consequently, due to its well-established status, it was seen as commonly labeled, i.e. signaled by the "abstract" label (see Figure 26). Conversely, when it comes to the EA, it was tentatively seen as encapsulating the content of an academic research article and providing additional information about it due to the co-occurrence of the BAxC with other constructions in its close vicinity. At the same time, it was also perceived as less established and thus not labeled. In other words, the EA, although soundly motivated by the general architecture of the constructicon (see Section 3.1.2.2.3.) and supported by the existence of related constructions, e.g. the video abstract (Spicer 2014), did not seem (yet) to come with a widely-recognized label. However, labels, being generally "relevant for representation of all kinds of concepts" and particularly useful for "categorization of elements that otherwise would be difficult to classify together", e.g. emerging discourse genre constructions, "work like other perceptual features such as shape, color, and size", at least at the beginning stages of development (Borghi and Binkofski 2014: 28). On this account, perceptual cues, such as those provided by grouping principles (see Section 3.1.1.1.2.1) could serve as "paralinguistic labels" (Strugielska 2022, private conversation), Such a "paralinguistic label", i.e. a separating white space, is illustrated in Figure 26. In practical terms, this observation meant that we were confronted with two types of corpus units and thus needed separate procedures to uncover the basic abstract construction and to discover the potential elaborated abstract construction.

5.1.1.1. The basic abstract construction

Due to the highly conventionalized, or emerged (see Section 2.3.) status of the basic abstract construction, its identification, inspired by the procedure developed for compiling ELFA (see Section 2.2.2.1.1.), was, roughly speaking, based on the linguistic label(s). In other words, we assumed that when the "abstract" label is provided, it should be seen as an indication that the members of academic ELF disciplinary communities identify a given token as a usage event pertaining to the BAxC.

However, while collecting the data, we realized that the tokens were not homogeneous. In fact, some of the tokens were labeled as "summary" (see Figure 27) even though their function could be easily identified as aligning with the BAxC, i.e. encapsulating the content of the article it precedes/follows.



Abstract

The aim of the paper is to identify the linguistic exponents of Russian compliments. The examples which will be analyzed come from contemporary Russian.

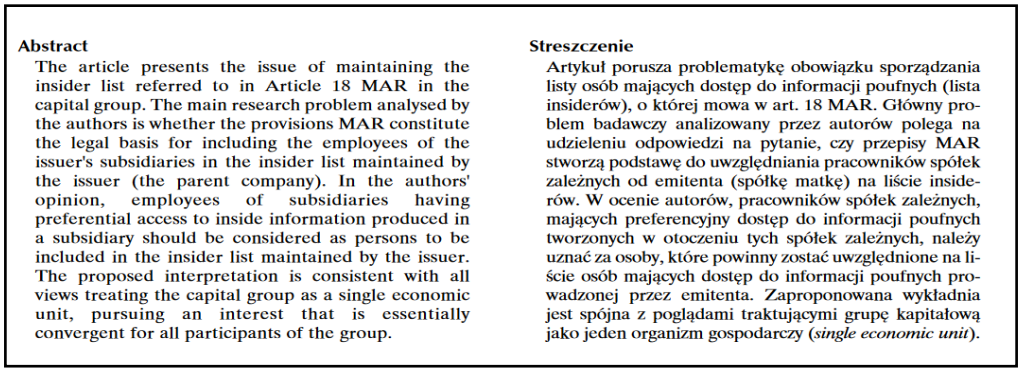
We will consider direct and indirect compliments, paying attention to such phenomena as presupposition and implicature as well as to the pragmatic functions of utterances. An analysis of communication strategies will allow us to present the specific features and role of compliments in linguistic communication in Russia.

Summary

The aim of the study described in this paper is to verify whether the structure of a modal predicate influences the type of modality expressed by the English modal verb *should*. The study uses language samples excerpted from *The corpus of contemporary American English*. It has adopted the model of *the semantic field of modal expressions* proposed by Angelika Kratzer. Additionally, this framework has been used to determine types of modality in this study. The analysis focuses on the interaction within the semantic field of the modal *should* with various forms of the main verb within the modal predicate structure.

Figure 27. "Abstract" and "summary" labels with BAxC

Hence, we decided to conduct a more detailed analysis of the tokens labeled "summary" and noticed that the label "summary" was treated by journal editors as equivalent to the "abstract" label. In particular, when a given journal provided tokens of abstracts in different languages, English tokens labeled as "abstracts" often co-occurred with Polish tokens labeled as "summaries" (i.e. streszczenia), while the content remained essentially the same[109] (see Figure 28).



Abstract

The article presents the issue of maintaining the insider list referred to in Article 18 MAR in the capital group. The main research problem analysed by the authors is whether the provisions MAR constitute the legal basis for including the employees of the issuer's subsidiaries in the insider list maintained by the issuer (the parent company). In the authors' opinion, employees of subsidiaries having preferential access to inside information produced in a subsidiary should be considered as persons to be included in the insider list maintained by the issuer. The proposed interpretation is consistent with all views treating the capital group as a single economic unit, pursuing an interest that is essentially convergent for all participants of the group.

Streszczenie

Artykuł porusza problematykę obowiązku sporządzania listy osób mających dostęp do informacji poufnych (lista insiderów), o której mowa w art. 18 MAR. Główny problem badawczy analizowany przez autorów polega na udzieleniu odpowiedzi na pytanie, czy przepisy MAR stworzą podstawę do uwzględniania pracowników spółek zależnych od emitenta (spółkę matkę) na liście insiderów. W ocenie autorów, pracowników spółek zależnych, mających preferencyjny dostęp do informacji poufnych tworzonych w otoczeniu tych spółek zależnych, należy uznać za osoby, które powinny zostać uwzględnione na liście osób mających dostęp do informacji poufnych prowadzonej przez emitenta. Zaproponowana wykładnia jest spójna z poglądami traktującymi grupę kapitałową jako jeden organizm gospodarczy (*single economic unit*).

Figure 28. "Abstract" and "streszczenie" labels with BAxC

Hence, since editors of ELF journals, i.e. members of academic ELF disciplinary communities producing publications for international purposes (see Section 2.1.1.) perceived the above-mentioned labels as equivalent, we decided to take into consideration the tokens labeled as both "abstract" and "summary"[110].

At the same time, we also found tokens classifiable as usage events pertaining to the BAxC which appeared without the "abstract" and "summary" labels. In this situation, we decided to rely on related websites, i.e. official websites attributed to journals through

[109] A similar practice was also noticeable when assessing PDF documents and their related websites. For instance, the text recognized on the website as an "abstract" (https://wuwr.pl/nkp/article/view/8202) was often positioned as a "summary" in the PDF file (e.g. https://wuwr.pl/nkp/article/view/8202/7835).
[110] In fact, as indicated in Section 4.2.2.2., the abstract construction is also referred to as summary.

which we had access to PDF files. In other words, if websites provided the label "abstract" or "summary" with reference to the same examples appearing in PDF files with no labels, such tokens were also taken into account[111] (see Figure 29).

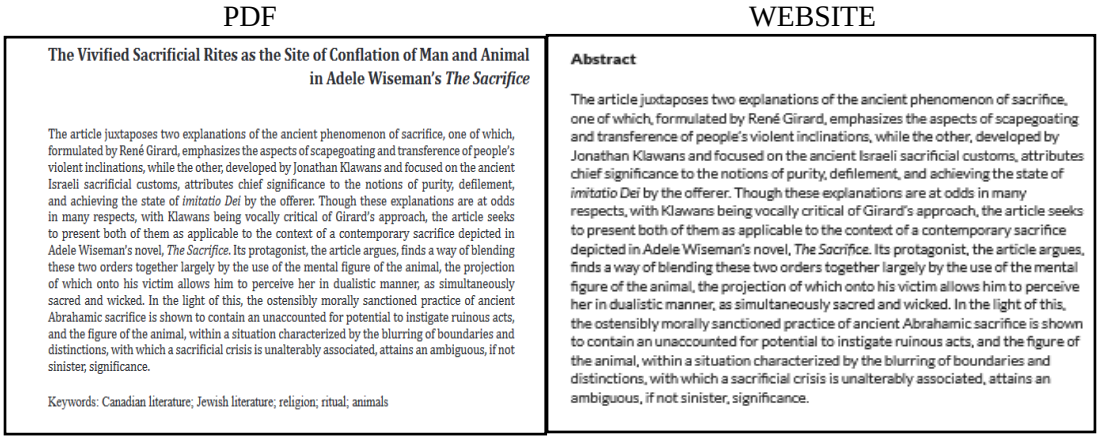


Figure 29. Difference between PDF files and websites

After identifying all tokens of the BAxC, we selected only those tokens which were within one page of PDF documents to facilitate our further work with the tokens of the potential elaborated abstract construction (see Section 5.1.1.2.2. for details). All in all, the procedure for identifying instances of the basic abstract construction can be depicted as follows:

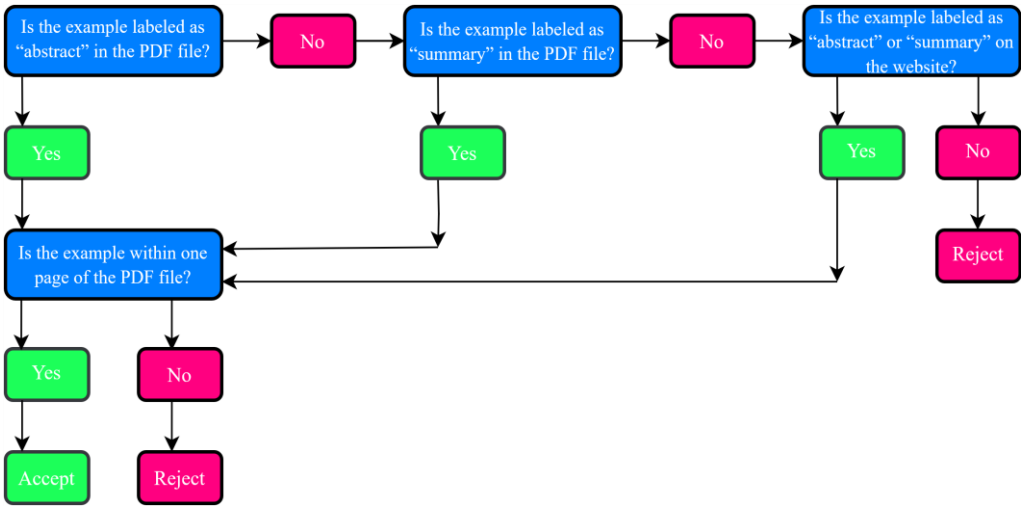


Figure 30. Procedure for identifying BAxC

[111] In this way, websites also served as the source of information indicating whether a given token is recognizable by the academic community as pertaining to the BAxC.

At this stage, all the accepted tokens were saved and functioned as the starting point for our main focus of analysis, i.e. the potential elaborated abstract construction.

5.1.1.2. Towards the elaborated abstract construction

When it comes to the EA, the identification of its tokens was a more challenging task. In fact, the procedure followed the route in the opposite direction from that developed for the identification of the BAxC. To be more specific, while in the case of the BAxC, a "pre-established construction" (see Section 2.2.1.1.) was our starting point, i.e. we took the basic abstract construction as already conventionalized, or emerged, which allowed us to concentrate on uncovering its tokens as well as noting details of its overall architecture, such as the distinction into labeled and labelless abstracts, in the case of the EA, we were initially guided by our intuitive judgments, which indicated a possible emergence of some kind of a new discourse genre construction, expanding and elaborating the scope of the BAxC. Since we could not help noticing that the basic abstract construction naturally groups with some of the other constructions within a PDF file , or, to paraphrase Mauranen and Sinclair (2006: 6), that grouping is "a natural and unavoidable way of perceiving" written data as they are "encountered" (see Section 2.2.1.1.), we decided to explore such groupings further.

In other words, once we confirmed that similar intuitions were shared by all members of the DISCOWER team, we decided to check if the groupings that we perceived recurred within the corpus. These steps, reflecting, in essence, the shift from emergent to emerging chunks proposed by ELF scholars working within LUG (see Section 2.2.2.2.1.), determined the way our work on the potential elaborated abstract construction was organized. To be more specific, the construction was discovered in two complementary stages, described in detail in Sections 5.1.1.2.1. and 5.1.1.2.2., which reflected the transition from the establishment of ad hoc, i.e. emergent, groupings by the DISCOWER team towards our attempts to capture their potential stability.

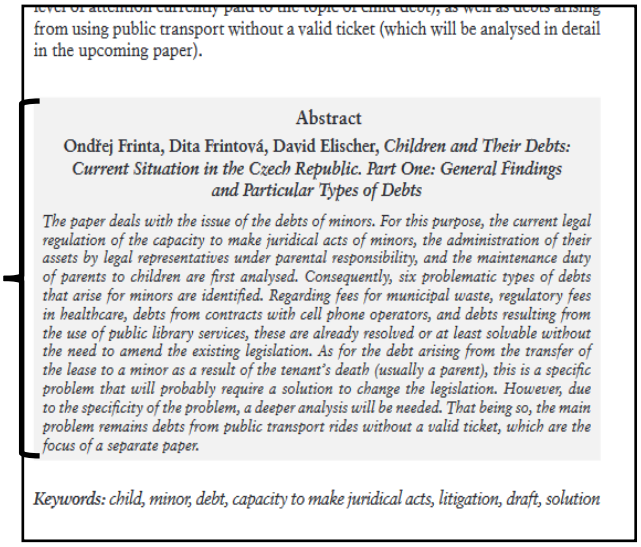
5.1.1.2.1. The emergent stage

As mentioned above, the emergent stage was associated with the delineation of ad hoc groupings by the team members, i.e. determining whether (and if so, what) groupings could be noticed in our data. At this stage, there was intentionally a lack of a well-defined procedure to guide this process, i.e. we basically segmented the PDF page around the tokens in an intuitive and spontaneous manner, much in the spirit of LUG.

Simultaneously, and also similarly to LUG, we realized that our intuitions resulted from an interplay of linguistic and paralinguistic cues, including paralinguistic labels, i.e. perceptual cues roughly corresponding to Gestalt principles of grouping (see Section 5.1.1.).

After individually analyzing each token, we discussed our "provisional unit boundaries"[112], roughly marking the beginnings and ends of the groupings established by each member of the DISCOWER team. All in all, the discussion indicated that there were three paralinguistic cues in our data guiding us in identifying tokens of the potential elaborated abstract construction. The first type, illustrated in Figure 31 below, was associated with a tinted background (see Hand and Middleditch 2014), i.e. a non-white background suggesting grouping by common region (see Section 3.1.1.1.2.1.).

Tinted background


from using public transport without a valid ticket (which will be analysed in detail in the upcoming paper).

Abstract

Ondřej Frinta, Dita Frintová, David Elischer, *Children and Their Debts: Current Situation in the Czech Republic. Part One: General Findings and Particular Types of Debts*

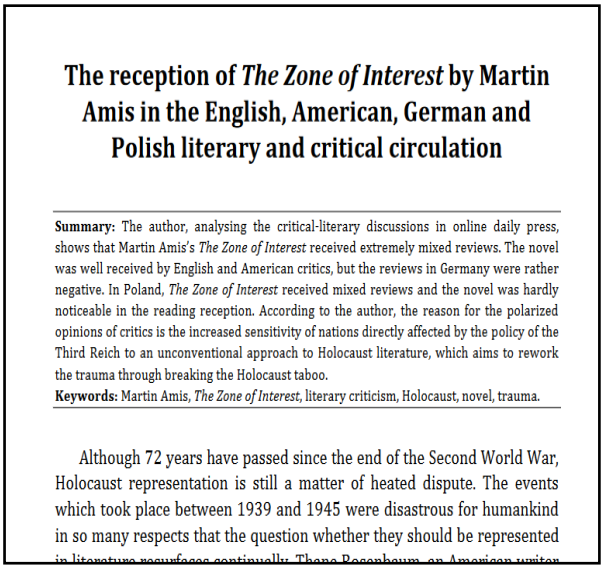
The paper deals with the issue of the debts of minors. For this purpose, the current legal regulation of the capacity to make juridical acts of minors, the administration of their assets by legal representatives under parental responsibility, and the maintenance duty of parents to children are first analysed. Consequently, six problematic types of debts that arise for minors are identified. Regarding fees for municipal waste, regulatory fees in healthcare, debts from contracts with cell phone operators, and debts resulting from the use of public library services, these are already resolved or at least solvable without the need to amend the existing legislation. As for the debt arising from the transfer of the lease to a minor as a result of the tenant's death (usually a parent), this is a specific problem that will probably require a solution to change the legislation. However, due to the specificity of the problem, a deeper analysis will be needed. That being so, the main problem remains debts from public transport rides without a valid ticket, which are the focus of a separate paper.

Keywords: child, minor, debt, capacity to make juridical acts, litigation, draft, solution

Figure 31. Tinted background

The second type referred to horizontal and/or vertical rules (see Waller 1980), i.e. (typically) black lines also suggesting grouping by common region (see Figure 32).

[112] As indicated in Chapter 2, the term was used by Mauranen and Sinclair (2006) to establish the boundaries of chunks in LUG.



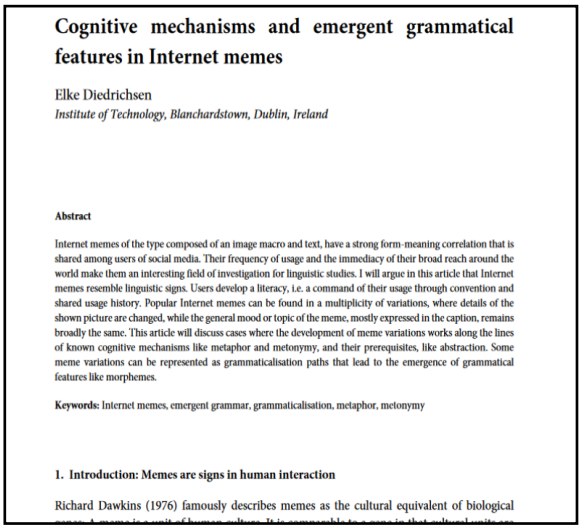
The reception of *The Zone of Interest* by Martin Amis in the English, American, German and Polish literary and critical circulation

Summary: The author, analysing the critical-literary discussions in online daily press, shows that Martin Amis's *The Zone of Interest* received extremely mixed reviews. The novel was well received by English and American critics, but the reviews in Germany were rather negative. In Poland, *The Zone of Interest* received mixed reviews and the novel was hardly noticeable in the reading reception. According to the author, the reason for the polarized opinions of critics is the increased sensitivity of nations directly affected by the policy of the Third Reich to an unconventional approach to Holocaust literature, which aims to rework the trauma through breaking the Holocaust taboo.
Keywords: Martin Amis, *The Zone of Interest*, literary criticism, Holocaust, novel, trauma.

Although 72 years have passed since the end of the Second World War, Holocaust representation is still a matter of heated dispute. The events which took place between 1939 and 1945 were disastrous for humankind in so many respects that the question whether they should be represented

Figure 32. Rules

The last type was related to white space (see Stöckl 2004), i.e. white gaps separating (typically) black objects on the page (see Figure 33).



Cognitive mechanisms and emergent grammatical features in Internet memes

Elke Diedrichsen
Institute of Technology, Blanchardstown, Dublin, Ireland

Abstract

Internet memes of the type composed of an image macro and text, have a strong form-meaning correlation that is shared among users of social media. Their frequency of usage and the immediacy of their broad reach around the world make them an interesting field of investigation for linguistic studies. I will argue in this article that Internet memes resemble linguistic signs. Users develop a literacy, i.e. a command of their usage through convention and shared usage history. Popular Internet memes can be found in a multiplicity of variations, where details of the shown picture are changed, while the general mood or topic of the meme, mostly expressed in the caption, remains broadly the same. This article will discuss cases where the development of meme variations works along the lines of known cognitive mechanisms like metaphor and metonymy, and their prerequisites, like abstraction. Some meme variations can be represented as grammaticalisation paths that lead to the emergence of grammatical features like morphemes.

Keywords: Internet memes, emergent grammar, grammaticalisation, metaphor, metonymy

1. Introduction: Memes are signs in human interaction

Richard Dawkins (1976) famously describes memes as the cultural equivalent of biological

Figure 33. White spaces

However, the identification of the three, or in fact two, grouping cues, i.e. common region and proximity, also revealed potential challenges. In fact, as predicted by Gestalt researchers (see Section 3.1.1.1.2.1.), we noticed that if one person was guided by the principle of common region and the other by the principle of proximity, divergent groupings emerged. If, however, we all followed the same principle, provisional group

boundaries were convergent[113]. This tendency, together with the commitment to develop a clear and repeatable methodology, prompted us to devise a procedure allowing for consistent identification of groupings.

In this way, we departed from LUG's dedication to intuition and pre-theoretical concepts, i.e. undefined chunks (see Section 2.2.1.1.), and moved towards a more objective means of deriving groupings. In fact, a similar solution, i.e. such that limits the role played by subjective factors in favor of a more rigid approach, is proposed by Mason (2008: 236), who finds LUG's "method of segmentation (using intuition) (…) unsatisfactory". In the same vein, Habjan (2013) suggests that the importance of intuition emphasized in LUG could gain a more scientific account if this intuition is fully explained in relation to the cognitive processes highlighted in usage-based approaches, e.g. Gestalt-based processes of grouping (see Section 1.2.2.)

5.1.1.2.2. The emerging stage

To develop our procedure for identifying tokens of the potential elaborated abstract construction, we first opted to concentrate on one Gestalt principle of grouping, i.e., proximity, seen as a "classic" and "primary" factor (see Section 3.1.1.1.2.1.). In fact, choosing proximity allowed us to depend on the indicator applicable to all our examples, as tinted boxes and rules, i.e. indicators of common region, constituted only a fraction of our dataset[114]. Consequently, we chose to compare white spaces between all (typically) black objects within PDF files to assess potential groupings.

Simultaneously, we decided to limit the scope of our analysis to one page of a PDF file to gather a maximally homogeneous collection of clear and (relatively) easily interpretable examples (see Section 2.2.2.2.1.). As a result, we were, for instance, able to display a page in a full-screen, single-page view and analyze the distances without the necessity to cross page boundaries, which, in fact, could hinder our analysis. A related obstacle that we managed to overcome by producing cleaned-up versions of the data, i.e. such that did not include incomplete elements, was the likelihood of being confronted with patterns not typical of and potentially not useful for "traditional linguistic analyses", such as C(C)G (Ungerer 2023: 9). In other words, we could much sooner than, for

[113] At this stage, we observed certain repetitions, e.g. the frequent co-occurrence of the basic abstract construction with its label (abstract/summary) and/or keywords (also often signaled by a label). Still, however, we were not aware of what the exact frequency of each chunk was and what the other stabilizing patterns could be.

[114] In other words, in every case, black objects within the PDF file were separated by means of white space.

instance, researchers working within the LUG framework, handle data that fairly closely resembled patterns which traditional grammars were in fact "designed to handle" (Mason 2008: 236). In essence, the procedure we developed can be summarized as follows:

1) Determine the largest gaps between black objects within the PDF file

Display the page in a full-screen, single-page view. Scan the page vertically and examine any white spaces that stretch horizontally across the entire page. This analysis encompasses both top and bottom margins, as well as any other white gaps between black objects. Find the largest gaps possible which suggest a grouping containing a token of the BAxC and of at least one different construction.

2) Determine whether all objects placed between the largest gaps on the page are complete

Check if the largest grouping contains black objects that do not fit within a single page. If not, such a cluster constitutes a token of the EA. If yes, exclude those objects and all those located above or beyond it closer to the page edge (see Figure 34). In this case, determine the largest gaps again, excluding the discarded objects.

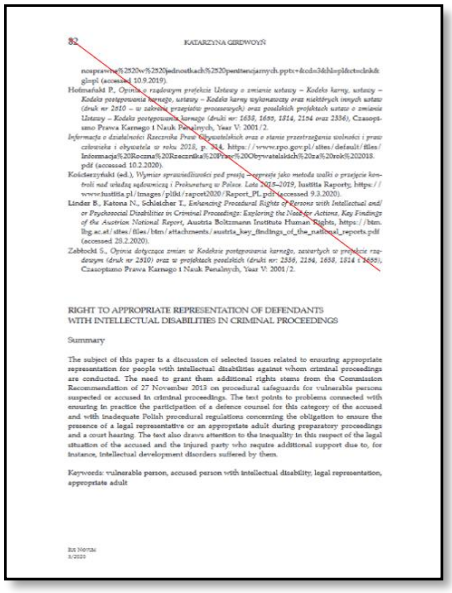


Figure 34. Black objects not within one PDF page

3) Determine all linguistic constructions contained within the delineated grouping

Write down which linguistic constructions (see Section 5.1.4.1. for details) are found in the grouping with the BAxC. Determine if there are smaller separating white spaces

between them. If yes, include this information by adding the separator "||" to indicate vertical separation and "- -" to indicate horizontal separation between the linguistic constructions. If not, use the separator "|" to indicate no separation between vertically arranged linguistic constructions and "-" to indicate no separation between horizontally arranged linguistic constructions.

To assess the reliability of our method for identifying the tokens of the EA (see Section 2.2.2.2.1. for the rationale), we calculated[115] our inter-rater agreement (see Section 2.2.2.2.1.). Each member of the DISCOWER team was asked to determine whether randomly selected objects (marked with dots in different colors) should be included in the largest grouping with the BAxC within the PDF file (see Figure 35).


Harnessing the Concept of an Array in Swift Programming Language. Abstract Concepts vs Natural Semantic Metalanguage

Bartłomiej Biegajło

Siedlce University of Natural Sciences and Humanities

Abstract

The contemporary digital world we all inhabit can have a twofold distinction; the majority of people participate in it in a passive manner; however, there is a growing number of IT enthusiasts who are keen to assume a more active role in the digitalized world society. The latter group of people want to take an active role not only in utilizing the digital inventions but, more importantly, in creating the digital world. This cannot be accomplished without developing a set of basic programming skills. To learn how to program applications requires perseverance and extended practice. This is especially problematic with regards to beginners who often feel overwhelmed by the number of keywords, programming patterns and various types of 'good practices' that they are advised to follow. Many of them quit prematurely as they suffer from the spreading preconception implying that programming is too exclusive. The intention of the following study is to test whether Natural Semantic Metalanguage could possibly help in more immediate accommodation of the newly acquired programming knowledge. Specifically, it is focused on developing an NSM explication of one of the core programming concepts, called an array, and possibly estimating the future potential behind such research.

KAREN GLASER

Universität Leipzig

karen.glaser@uni-leipzig.de

ORCID: 0000-0002-2527-7415

Enhancing the role of pragmatics in primary English teacher training

ABSTRACT. Teaching EFL in primary school is no longer a novelty but firmly established in the education landscape throughout Europe and many countries worldwide. Primary English language teaching (PELT) is a unique branch of ELT insofar as it entails both the teaching of children and beginners. While PELT teachers and PELT teacher educators largely agree that this concurrence of 'young plus beginning' requires a focus on vocabulary, speaking and listening, introduced and practiced through songs, games, stories, roleplaying and embodiment techniques such as Total Physical Response, pragmatic aspects often take a backseat in PELT teacher training and by extension in the PELT classroom, even though it has been established that pragmatics instruction is necessary and feasible on all proficiency levels, right from the beginning. This article discusses possible reasons for this omission and illustrates with authentic examples why pragmatics should play a bigger role in the training of primary English teachers.

KEYWORDS: EFL, English Language Teaching, pragmatics, pragmatic awareness, language teacher training, primary school, Primary English Language Teaching (PELT).

1. INTRODUCTION

The past two decades have seen the widespread introduction of early foreign language teaching throughout Europe and other parts of the world, with most school curricula making the teaching of at least one foreign language compulsory at primary level (Edelenbos, Johnstone & Kubanek 2006). Due to the heterogeneous linguistic landscape within the EU, there is a great deal of variation as to the chosen target language(s), but despite all local variation, English has transpired as the most dominant target language. This

Figure 35. Examples for evaluating inter-rater agreement

With four elements under consideration in each example, we categorized the examples into 16 potential groups, and we assumed that each category could occur with equal probability. In other words, each example could be evaluated in 16 different ways by the team members. The evaluation of examples was conducted independently under the same conditions, i.e. each team member assessed the examples on the same screen. The evaluation process lasted less than 90 minutes. This process was repeated for 87 examples randomly chosen from the corpus and the resulting Randolph's free-marginal interrater

[115] We relied on the online kappa calculator available at http://justusrandolph.net/kappa/

kappa value was 0.843[116], indicating an "excellent" level of agreement in Fleiss's (1971) interpretation[117]. Subsequently, each team member determined the (lack of) elaboration for each token of the BAxC in the corpus, determining the boundaries of each elaborated form. Next, the results were compared to determine agreement on which tokens in the corpus could receive descriptions regarding their elaborated form. The discussion revealed that our boundaries were, expectedly, mostly convergent. In cases where there was disagreement among the team members, we opted for groupings determined by the majority.

5.1.2. External criteria

Concurrently, the choice of tokens included in the corpus was influenced by several external criteria, which can be divided into two types, i.e. those related to abstracts and those related to disciplinary communities. The former group encompasses factors impacting the selection of tokens of the basic abstract construction (Section 5.1.2.1). The latter, in turn, refers to factors regarding the selection of academic disciplines from which we extracted our tokens of the BAxC (Section 5.1.2.2.).

5.1.2.1. Criteria related to abstracts

Firstly, we took into account only the tokens of the BAxC available in open-access journals in PDF files. By concentrating on the data in open-access journals, we expedited the data collection process as we assumed that each author had given consent for the utilization of their content. By concentrating on the data in PDF files, we ensured a relatively stable description of corpus units. To clarify, while many websites included the tokens of the BAxC, they underwent frequent updates and relocations, influencing the co-occurrence of the basic abstract construction with other constructions in its close vicinity. Thus, incorporating the tokens derived from websites into the corpus could lead to inconsistent descriptions due to the ongoing changes.

Secondly, we included only the tokens of the basic abstract construction from PDF files in which articles were written in English and there was no explicit indication concerning the translation by a specialist. This step allowed us to ensure that each token

[116] This value can be extended over the whole corpus with a 95% confidence level and a 5% margin of error.

of the basic abstract construction included in the corpus was written by an academic ELF user, thereby preserving the integrity of the data.

5.1.2.2. Criteria related to disciplinary communities

When it comes to criteria pertaining to the selection of disciplinary ELF communities, we started from the assumption that we want to collect data representing ELF users. Hence, in accordance with the definition of academic ELF provided in Section 2.1.2., both native and non-native authors, as well as other members of the academic community contributing to their authorship, publishing in English for international purposes were taken into account.

Next, we decided to focus specifically on abstracts from Polish journals. This approach was aimed at ensuring that abstracts appearing in English in peripheral (see Section 2.1.2.), or local, journals were included. In this way, we wanted to guarantee that both the authors of basic, and thus also elaborated, abstracts as well as, e.g., their editors, were ELF users.

When selecting disciplines, (see Section 2.1.1.), we opted for an approach regarded as the most pragmatic, i.e. reliance on official (inter)national lists classifying academic disciplines. Due to our focus on Polish journals, we decided to utilize a list of journals compiled by the Ministry of Science and Higher Education in which journals were officially assigned to a given discipline. Consequently, by using such documents and institutional policies as our basis for defining the disciplines and journals under study, we established the beginning of the timeframe we were interested in. More specifically, we decided to collect data from 2018 (in which Poland adopted a new disciplinary classification) to 2021 (in which our project started).

Out of all the disciplines, we focused on three, i.e. law, linguistics, and literary studies. Given that WrELFA operates within a very broad division between sciences and social sciences and humanities (see Section 2.2.2.1.2.), we decided to concentrate on a more detailed subdivision within the latter category. More specifically, being a discipline closely related to the research interests of the team members, linguistics served as our starting point. To identify the remaining disciplines, we chose to include disciplines from social sciences and humanities which, based on the classification we followed, could be identified as strongly and weakly related to linguistics.

To measure this relationship, we used the above-mentioned list and counted the occurrences of Polish multidisciplinary journals that included both linguistics and another

discipline in their profile. The analysis indicated a strong association between linguistics and literary studies and a weak one between linguistics and law. Therefore, our corpus contains examples from two disciplines representing humanities, namely linguistics and literary studies, and one representing social sciences, namely law. Details regarding the quantitative representation of data contained within DISCOWER are provided in the following section.

5.1.3. Balance

The DISCOWER corpus includes tokens collected from 87 journals, i.e. 33 from law, 33 from linguistics, and 21 from literary studies. When it comes to the BAxC, there are 2136 tokens, i.e. 987 from law, 819 from linguistics, and 330 from literary studies. Hence, while the data concerning law and linguistics can be perceived as quite balanced, the balance between linguistics/law and literary studies is not achieved. However, when the distinction between humanities and social sciences is considered, the corpus appears to be reasonably well-balanced (see Figure 36).

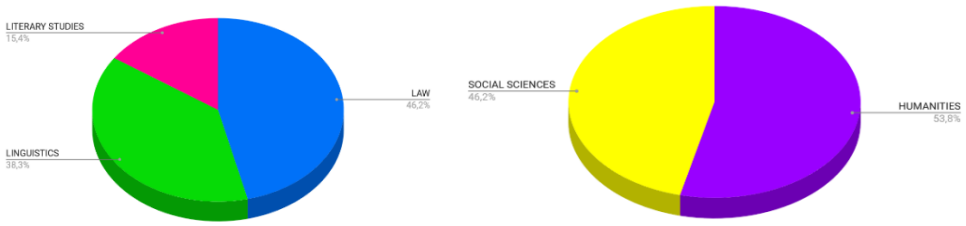


Figure 36. BAxC: quantitative data

At the same time, there are 2108 tokens of the EA, i.e. 966 from law, 817 from linguistics, and 325 from literary studies. On this account, since in only 28 cases it was not possible to delineate tokens of the EA, the interpretation of corpus balance is similar, i.e. the data regarding law and linguistics as well as the data concerning social sciences and humanities may be considered relatively even (see Figure 37).

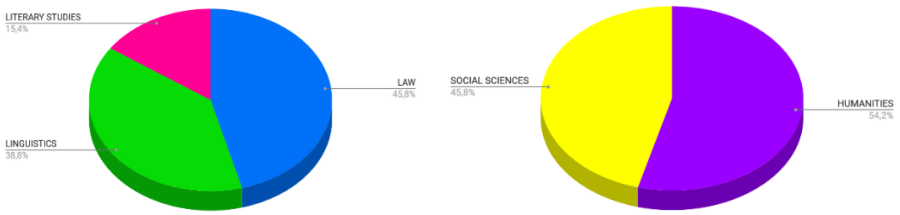


Figure 37. EA: quantitative data

Apart from the classification based on disciplines, our data appear quite balanced when considering, among others, the years of publication. To clarify, for the BAxC, we have 594 tokens from 2018, 699 from 2019, 603 from 2020, and 240 from 2018. For EA, we have 585 tokens from 2018, 686 from 2019, 599 from 2020, and 238 from 2018. Hence, considering both types of data, a reasonably balanced proportion is achieved with reference to 2018, 2019, and 2020 (see Figure 38).

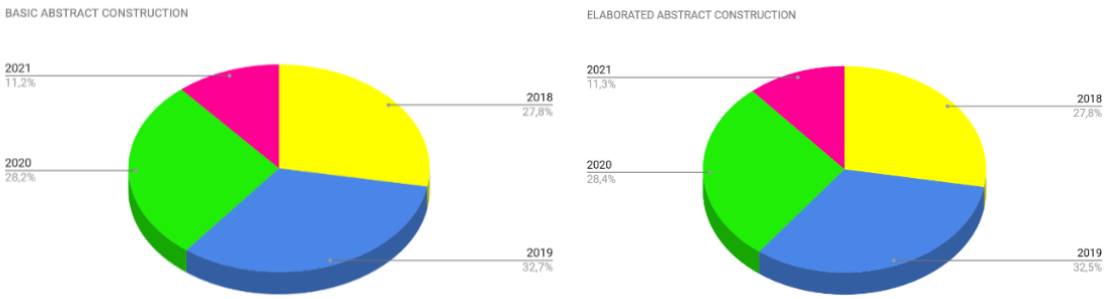


Figure 38. Years of publication: quantitative data

However, considering the countries of affiliation, the corpus, expectedly, appears to be highly skewed towards academic ELF users from Poland. More specifically, the corpus contains data written by users from 76 countries, among which the top 10 most represented countries are: Poland (basic: 1253, elaborated: 1235), the United States of America (basic: 100, elaborated: 96), Italy (basic: 68, elaborated: 68), Germany (basic: 61, elaborated: 61), Spain (basic: 56, elaborated: 56), Czechia (basic: 52, elaborated: 52), the United Kingdom (basic: 42, elaborated: 40), Hungary (basic: 37, elaborated: 37), Ukraine (basic: 31, elaborated: 31), and Iran (basic: 31, elaborated: 31)[118]. Still, apart from Poland, a fairly even proportion among the remaining 9 countries included in the corpus is achieved (see Figure 39).

[118] If a given token was written by multiple authors from, for example, Poland, it was counted as a single token influencing the number of authors from Poland. However, if a given token was written by authors from, for example, Poland and Italy, it was counted as a token influencing the number of authors from both Poland and Italy.

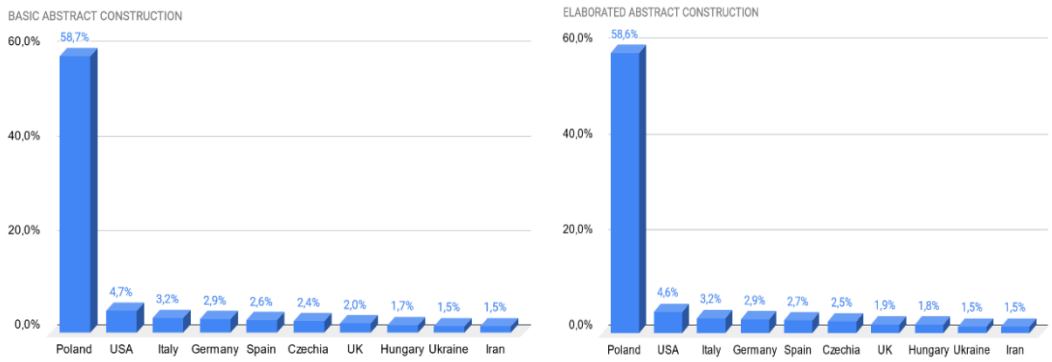


Figure 39. 10 most represented countries in DISCOWER

5.1.4. Data description

The corpus data are available in two forms. Firstly, users can visit the corpus browser (https://serwisy.umk.pl/discower/), where the data can be searched based on various criteria, e.g. discipline, publication year, journal, or the (lack of) labels attributed to the basic abstract construction. Secondly, a downloadable spreadsheet is available, containing 30 columns mirroring the aforementioned aspects. At the same time, both in the browser and the spreadsheet, data can be searched based on information related to linguistic and/or paralinguistic constructions.

More precisely, given the increasing interest in the multimodal nature of discourse genre constructions (see Section 4.2.) and the consequent need for further research deepening the relationships between linguistic and paralinguistic constructions (see Section 4.3.), we decided to include in the corpus data that enable further exploration of the BAxC and the EA by means of the aforementioned constructions. However, given our main focus on the EA, data pertaining to both linguistic and paralinguistic constructions are available at this stage of corpus development only in relation to the EA[119]. Therefore, the following sub-sections focus on a detailed description of the data concerning the potential elaborated abstract construction, which, in fact, is the main subject of investigation in Chapter 6.

[119] When it comes to the description of the basic abstract construction, at the current stage of corpus development, data concerning paralinguistic constructions are available. These data are structured in a way that allows for potential exploration of the part-whole relationship, i.e. between the base and its elaboration, relying on corresponding data. For instance, at both levels, it is possible to examine the continuity construction.

5.1.4.1. Linguistic constructions

Considering our reliance on the cognitive and constructionist commitments developed in C(C)G, our starting point for determining and describing linguistic constructions was to evoke a reference point, i.e. a label, through which its dominion, including the target, i.e. the label's content, could be accessed (see Section 3.1.2.2.3.). In other words, building on our experience in corpus creation, specifically on using abstract labels to identify tokens of the BAxC, we assumed that the identification and description of other linguistic constructions should also be feasible using a similar approach.

Therefore, building on solutions employed in researching academic ELF (see Section 2.2.2.2.2.) and discourse genre constructions (see Section 4.2.1.2.), we decided to first develop a set of constructions derived from explicit norms (or, in other words, determine the emerged constructions). To this end, we sought inspiration from both publications describing abstracts and their co-occurrence with other discourse genre constructions. Initially, we were aware that the basic abstract construction can be surrounded by, e.g., labels, keywords, or titles (see Figure 35). Such information, as noted by Nasar et al. (2018), holds great importance for research articles because "researchers tend to decide paper relevance with their domain of interest based on metadata information such as title, abstract, references, authors, citing articles and affiliations". Consequently, such metadata information is often extracted from documents to create databases facilitating publication searches[120] and is thus categorized in various ways. Therefore, our literature review focused on such categorizations (Seymore et al. 1999; Nasar et al. 2018), by means of which we were able to establish that the basic abstract construction can be surrounded by information related to, among others, titles, authors (and their affiliations or e-mails), DOI addresses, publication dates, notes, journals, keywords, or pages (see Figure 40) and identify their function[121].

[120] In other words, many research organizations establish digital research repositories to offer search filters based on such metadata to make it easier for users to find research articles.

[121] Linguistic constructions are also potentially describable in terms of Gestalt schematic structures. Since however, linguistic constructions are specific constructions and thus rest primarily on rich knowledge structures and only secondarily on schematic ones (see Section 3.1.1.3.1), the link between linguistic constructions and Gestalt schematic structures is intentionally neglected.

Figure 40. Metadata concerning BAxC

Secondly, to be consistent with our previous approach to recognizing tokens of the basic abstract construction based on labels, we searched for words or phrases in our data which were separated by punctuation marks or distinguished, among others, by appearing in bold. Admittedly, the labels we found mostly complied with the information derived from the relevant literature, as the labels referred to, for instance, e-mails, DOI addresses, or dates of publications (see Figure 41).

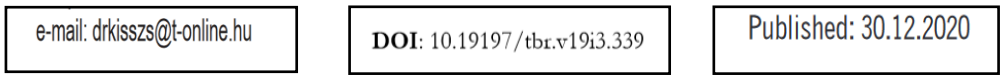


Figure 41. Examples of labels in DISCOWER

Hence, the set of constructions was divided into two types, i.e. constructions which appeared with their labels and those which did not. In the first case, to signal such a relationship, we divided the relevant pairs into labels and contents. For instance, we referred to tokens of the BAxC as “abstract (content)” and to their labels as “abstract

(labels)"[122]. In the second case, the distinction into labels and contents was not applied and we relied only on the name we assigned to a given construction, e.g. article (title).

Furthermore, building on the attribute-value relationship underlying the semantic pole of constructions (see Section 3.1.1.3.2.), we divided the constructions into several groups, i.e. attributes, with multiple subgroups, i.e. values. Selected examples based on such a relationship are provided below.

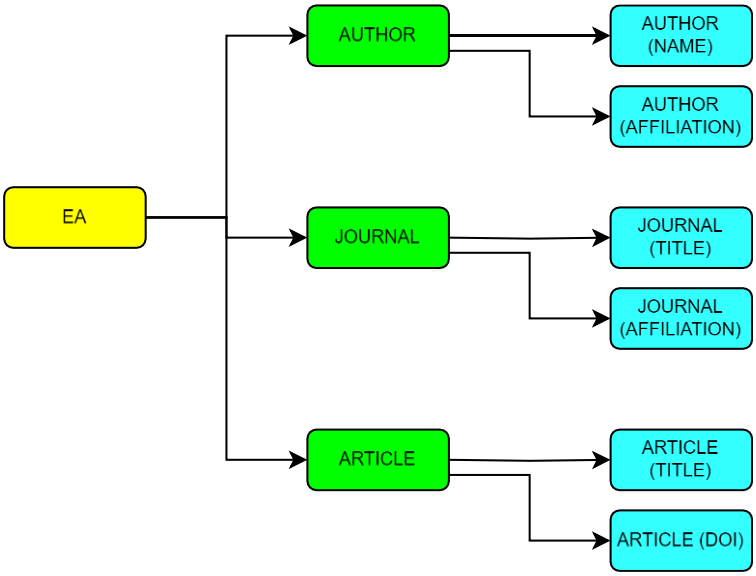
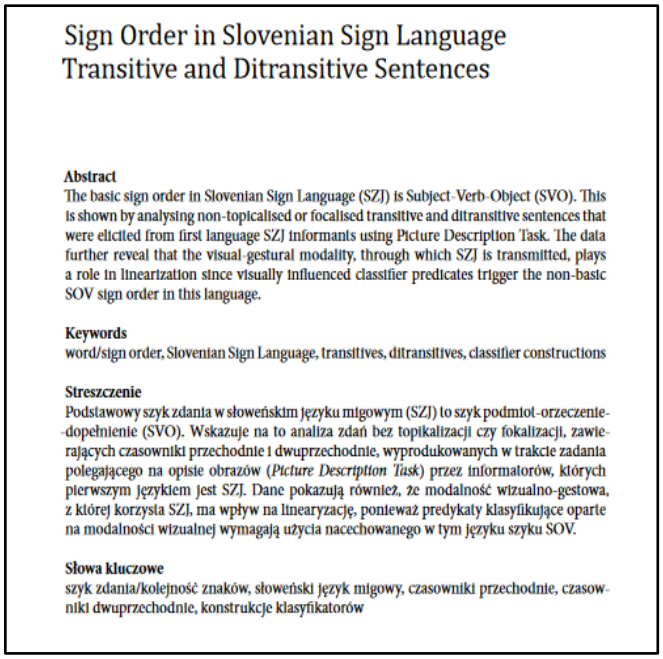


Figure 42. Attribute-value relationship (linguistic constructions)

Concurrently, if a given token of the potential elaborated abstract construction contained tokens of constructions in languages different than English, this information was also provided. Hence, a given token of the potential elaborated constructions could be described with reference to such discourse genre constructions, as, for instance, article (title), abstract (label), abstract (content), keywords (label), keywords (content), abstract (label: summary) (Polish), abstract (content) (Polish), keywords (label) (Polish), and keywords (content) (Polish) (see Figure 43).

[122] However, we also referred to- tokens of the basic abstract construction as abstract (content) in cases in which the abstract label was either non-present in PDF files or not included in the scope of the elaboration. In other words, while not every token of the X construction was explicitly labeled as such, the fact that in some cases they were signaled by the X label made us assign the name X (content) to all the tokens, irrespective of their co-occurrence with their related labels.



Sign Order in Slovenian Sign Language
Transitive and Ditransitive Sentences

Abstract
The basic sign order in Slovenian Sign Language (SZJ) is Subject-Verb-Object (SVO). This is shown by analysing non-topicalised or focalised transitive and ditransitive sentences that were elicited from first language SZJ informants using Picture Description Task. The data further reveal that the visual-gestural modality, through which SZJ is transmitted, plays a role in linearization since visually influenced classifier predicates trigger the non-basic SOV sign order in this language.

Keywords
word/sign order, Slovenian Sign Language, transitives, ditransitives, classifier constructions

Streszczenie
Podstawowy szyk zdania w słoweńskim języku migowym (SZJ) to szyk podmiot-orzeczenie-dopełnienie (SVO). Wskazuje na to analiza zdań bez topikalizacji czy fokalizacji, zawierających czasowniki przechodnie i dwuprzechodnie, wyprodukowanych w trakcie zadania polegającego na opisie obrazów (*Picture Description Task*) przez informatorów, których pierwszym językiem jest SZJ. Dane pokazują również, że modalność wizualno-gestowa, z której korzysta SZJ, ma wpływ na linearyzację, ponieważ predykaty klasyfikujące oparte na modalności wizualnej wymagają użycia nacechowanego w tym języku szyku SOV.

Słowa kluczowe
szyk zdania/kolejność znaków, słoweński język migowy, czasowniki przechodnie, czasowniki dwuprzechodnie, konstrukcje klasyfikatorów

Figure 43. Constructions in different languages

All in all, we established 86 linguistic constructions by means of which the potential elaborated abstract construction can be described, among which, apart from abstract (content), the top ten are: keywords (label), keywords (content), abstract (label), article (title), author (name), author (affiliation), author (contact) (content), abstract (label: summary), article (DOI) (content), keywords (label) (Polish). The illustration of their forms and meanings is provided in Table 1 below.

CONSTRUCTION	GRAPHOLOGICAL POLE	SEMANTIC POLE
keywords (label)	the word keywords	identifying the following content as keywords
keywords (content)	a set of words and/or phrases separated by punctuation marks	indicating the main topic(s) addressed in an article
abstract (label)	the word abstract	identifying the following content as an abstract
article (title)	a phrase and/or a clause	indicating the main title attributed to an article
author (name)	proper nouns	indicating the name and surname of an author
author (affiliation)	nouns and/or proper nouns	indicating a university where an author works or studies
author (contact) (content)	a combination of letters and words forming an e-mail address	indicating an email address through which one can contact an author

abstract (label: summary)	the word summary	identifying the following content as a summary
article (DOI) (content)	a combination of letters and words forming a URL address	indicating a URL address serving as a persistent identifier of an article
keywords (label) (Polish)	the phrase słowa kluczowe	identifying the following content as słowa kluczowe

Table 1. 10 most frequently appearing linguistic constructions in DISCOWER

Based on such data, the potential elaborated abstract construction can be examined in relation to typical clines attributed to constructions (see Section 3.1.2.2.). In other words, these data can be searched to determine which sequence is more frequent, e.g. abstract (label) - abstract (content) | keywords (label) - keywords (content) or abstract (label) - abstract (content), thereby allowing for the description of the EA considering varying degrees of fixedness. Additionally, due to various combinations of abstract (content) with abstract (label: summary), abstract (label: summary) (Polish), or abstract (label), it is possible to establish a more schematic construction related to the connection of abstract (content) with abstract (label_X). Simultaneously, concerning different degrees of complexity, such constructions can also be investigated regarding the number of identifiable linguistic units, exploring the potential elaborated abstract construction in terms of sequences comprising, for example, 4 or 13 constructions.

5.1.4.2. Paralinguistic constructions

When it comes to paralinguistic constructions, our main reference point was the cognitive operation of grouping (see Section 3.1.1.1.2.1.) and, consequently, reification (see Section 3.1.1.1.2.2.). More specifically, based on the assumption that identifying the potential elaborated abstract construction resulted from a grouping process guided by Gestalt processes, we assumed that the resulting group should be subsequently described in terms of relevant features, i.e. paralinguistic constructions based in Gestalt schematic structures (see Section 3.1.2.2.2.).

To achieve this, we firstly decided that such constructions can be categorized as relating to more intrinsic and more extrinsic features of the potential elaborated abstract construction (see Section 3.1.1.3.1.), thereby focusing on two main Gestalt principles and their corresponding Gestalt schematic structures, i.e. closure and similarity. Therefore, the description of the potential elaborated abstract construction can be approached

through constructions that relate to its shape[123] (intrinsic) and/or similarity/dissimilarity to other constructions on the page or among the ones composing the EA (extrinsic).

As a result, paralinguistic constructions were also grouped based on the attribute-value relationship. In this case, however, the values were divided in a binary manner, or digital (Section 3.1.1.3.2.), i.e. whether the potential elaborated abstract construction possesses a feature aligning with maximal goodness or not (see Section 3.1.1.1.2.2.). Selected examples are illustrated in Figure 44 below.

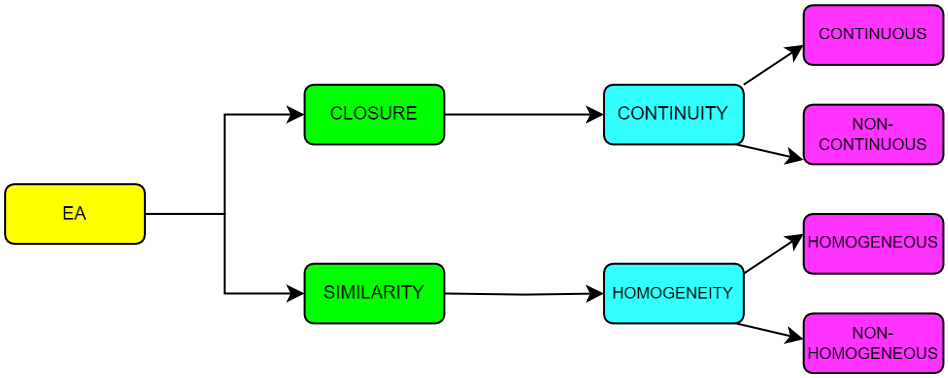


Figure 44. Attribute-value relationship (paralinguistic constructions)

Homogeneity, for instance, is based on the assumption that maximal goodness entails maximal similarity. Thus, any deviations through punctuation within the potential elaborated abstract construction, such as bold or italics, are perceived as deviations from maximal goodness, meaning that the potential elaborated abstract construction can be described with reference to the value non-homogeneous.

Continuity, in turn, pertains to the shape of the EA, which, considering its recognized significance in the context of discourse genre constructions (see Section 4.2.2.1.), is the main focus of the study described in Chapter 6, analyzing the co-occurrence of linguistic and paralinguistic (shape-related) constructions. Therefore, in the following sections, methods of operationalizing the shape of the potential elaborated abstract construction and related paralinguistic constructions are presented.

[123] The notion of shape is applied in the present chapter to indicate that it is used in relation to the (intrinsic) content structure elaborating the schematic structure (contour) (see Section 3.1.1.3.1). In other words, it is used with reference to a contour of a specific object, i.e. abstract, and is thus named as its shape.

5.1.4.2.1. Shape

To determine the degree of overlap with the maximally good shape, we divided the evaluation stage into two phases. Firstly, we identified the outermost black fragments of each token of the potential elaborated abstract construction and marked along these fragments a relatively adjacent line showing the boundary between a black and a white fragment of a PDF page[124]. At this stage, we aimed to assess the edges of the tokens in the most objective manner and thus we did not add any missing elements to create an ideal (i.e. gapless) shape (see Figure 45).

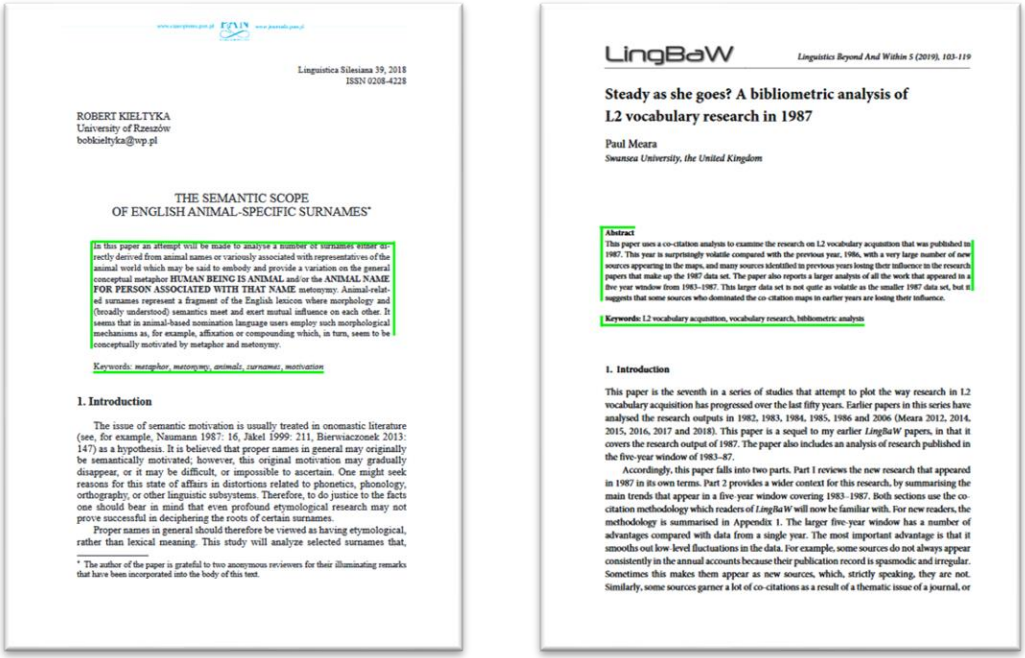


Figure 45. Shape delineation

As shown in Figure 45, a straight line was added as long as there was no difference in interline spacing. If such a difference occurred, the line was not drawn, thus creating a gap in the shape.

Secondly, based on the outlined edges of each token, the potential maximum extent of black objects was determined by drawing a line without any gaps to imitate the ideal realization. Next, the real edges, i.e. those visible on the page, and the ideal edges

[124] In fact, this method roughly aligns with the approach adopted in research on the visual segmentation of pages incorporating Gestalt principles. In other words, edges of objects within a page are detected when there is a noticeable shift in, among others, color (see, for instance, Dou and Kong 2008).

were compared. Places where the two versions did not align were marked in red to identify the missing fragments (see Figure 46).

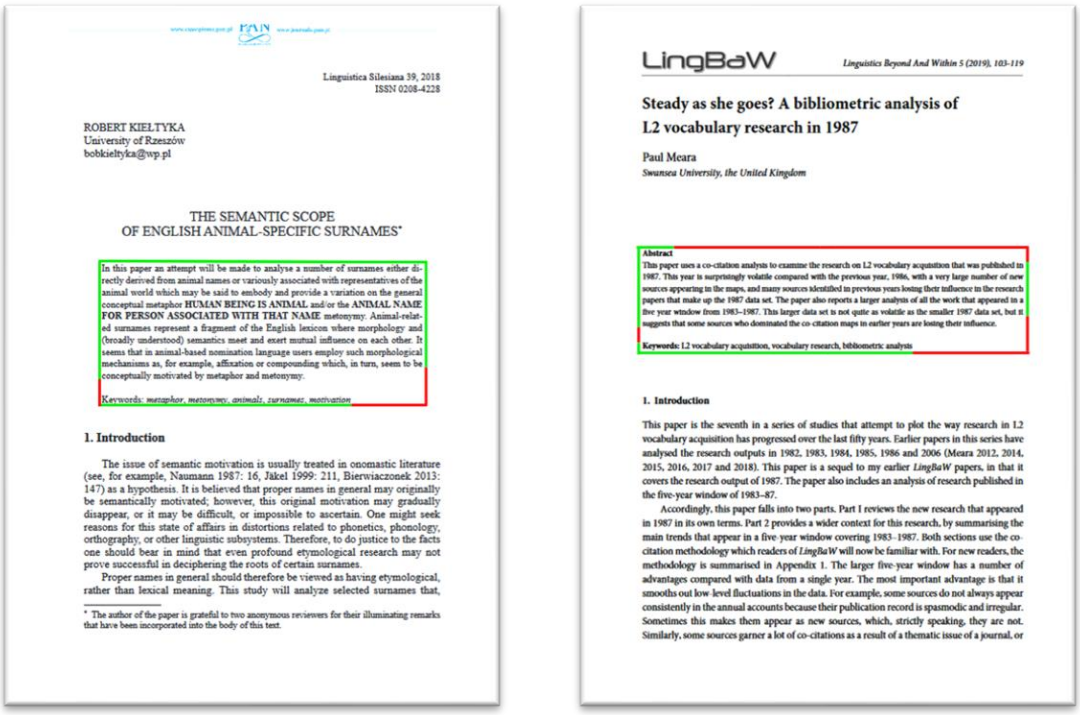


Figure 46. Differences between real and ideal shapes

However, the analysis of each token revealed that the tokens included in the corpus are highly diverse. In other words, the edges could either lack gaps or have them, which, at the same time, could be located in different positions, e.g. closer to the left or right boundary of the page. Therefore, to establish a common ground for 2108 tokens in the corpus, we decided to narrow down the description of shape to selected paralinguistic constructions.

5.1.4.2.2. Shape-related paralinguistic constructions

In fact, as noted in Section 3.1.1.1.2.2., the influence of embeddedness on shapes cannot be overlooked; therefore, to evaluate the shape of the EA and shape-related constructions properly, we decided to identify prominent fragments of shapes related to objects that are read on screens.

As noted by Mattice (2020: 221), “when we read on screens, we tend to read in a Z or F pattern”. The Z pattern is associated with the Gutenberg principle, which states that a page can be divided into four quadrants, i.e. the primary optical area, the terminal

area, the strong fallow area, and the weak fallow area (see Figure 47). According to this principle, in Western cultures, readers' eyes go straight to the upper left hand corner, then look towards the bottom right" (Frick and Eyler-Werve 2014: 83). Hence, while fallow areas are said to receive attention only when they are made prominent on purpose (Lidwell and Holden 2010), the upper left corner and the bottom right corner attract the greatest attention.

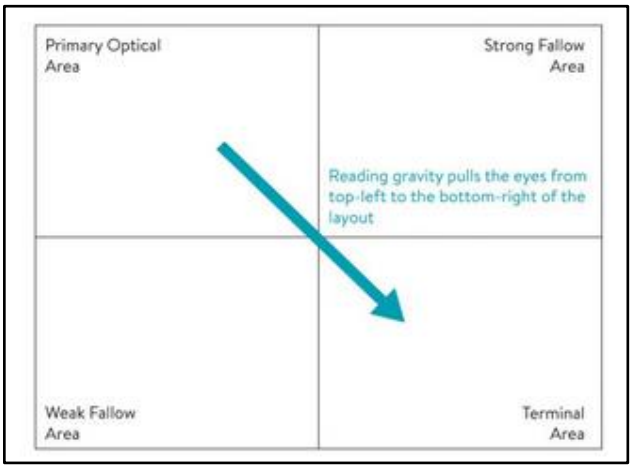


Figure 47. Gutenberg principle (Frick and Eyler-Werve 2014: 84)

However, as Lidwell and Holden note (2010: 118), the Z pattern is "likely only predictive of eye movements for heavy text information, evenly distributed and homogeneous information, and blank pages or displays".

The F pattern, in turn, is discussed with reference to online texts (Nielsen 2006). As Nielsen (2006) explains, users start by reading horizontally across the upper part of the content, then move down slightly and read horizontally again, and finally scan vertically along the left side of the content (see Figure 48).

Figure 48. Heatmaps (Nielsen 2006)

Hence, not wanting to distinguish between heavy and non-heavy text information[125], we took into account the results regarding the two aforementioned reading patterns and focused on describing the upper left corner, the bottom right corner, as well as the left edge of the tokens.

In line with the previous approach involving the development of a classification for linguistic constructions, we decided to establish explicit norms regarding paralinguistic constructions (Waller 1980; Tversky 2001; Stöckl 2004; Patt 2013). For example, theories concerning macropunctuation were used to determine the equivalents of micropunctuation, e.g. commas or semi-colons, that can be seen as relevant for the evaluation of the PDF page. Hence, it was established that, for instance, white spaces or rules indicate "delineation". In other words, they are seen as "methods of indicating the beginning and end of text segments" (Waller 1980: 248). As a consequence, we assumed that white space functions broadly as the indicator of groupings or groupings within groupings. Next, it was discovered that, for example, indentation is equivalent to indicating subordinate relationships through paralinguistic constructions (Tversky 2001). Thus, we determined, for example, that a paralinguistic construction whose graphological pole entails a white space and whose semantic pole entails a grouping can be perceived as a general indicator of a punctuation unit within the PDF page. Therefore, we classified such a construction as relating to simplicity (see Figure 49). At the same time, we also established the construction independence, whose graphological pole entails indentation and whose semantic pole entails showing (in)dependence in relation to other elements within the PDF (see Figure 49).

Hence, simplicity was evaluated with reference to such values as simple and non-simple, which were determined based on the size(s) of white spaces. To clarify, if there was not a noticeable contrast in the size of white gaps, i.e. indicating homogeneous interline spacing, a given token was perceived as consisting of a single part (and thus simple). However, if such a difference was noticeable, a given token was perceived as non-simple. Currently, if such a separation was noticed, either horizontally or vertically, we added the separators, "||", "- -", to the description of sequences of linguistic

[125] In other words, tokens of the potential elaborated abstract construction were highly heterogeneous, i.e. the corpus contains tokens consisting of 2 or even 30 linguistic constructions. Therefore, without a clear indication of when a particular realization can be considered heavy, we did not introduce this distinction into DISCOWER.

constructions constituting the EA to delineate potential boundaries of punctuation units (see Section 5.1.1.2.2.).

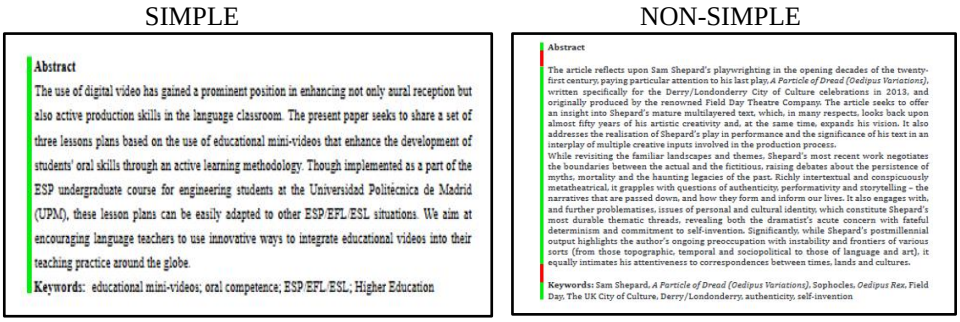


Figure 49. Simple and non-simple value

On this account, we developed a set of paralinguistic constructions whose graphological and semantic poles were connected to the previously identified essential fragments of objects read online (see Table 2).

FORM	MEANING
the upper left corner present	indicating independence
the upper left corner absent	indicating non-independence
the bottom right corner present	indicating completeness
the bottom right corner absent	indicating non-completeness
homogeneous alignment of parts present	indicating continuity
homogeneous alignment of parts absent	indicating non-continuity

Table 2. Paralinguistic constructions related to shape

In other words, we developed a set of shape-related paralinguistic constructions, i.e. independence, completeness, and continuity, and each category (i.e. attribute) was further divided into two possible values. To put it differently, the values were operationalized in a binary manner, i.e. a given token could be described as either possessing or non-possessing the value of an ideal shape.

To illustrate, when it comes to independence, a given token could be described as either independent or non-independent, depending on the presence of the upper left corner (see Figure 50).

INDEPENDENT DEPENDENT

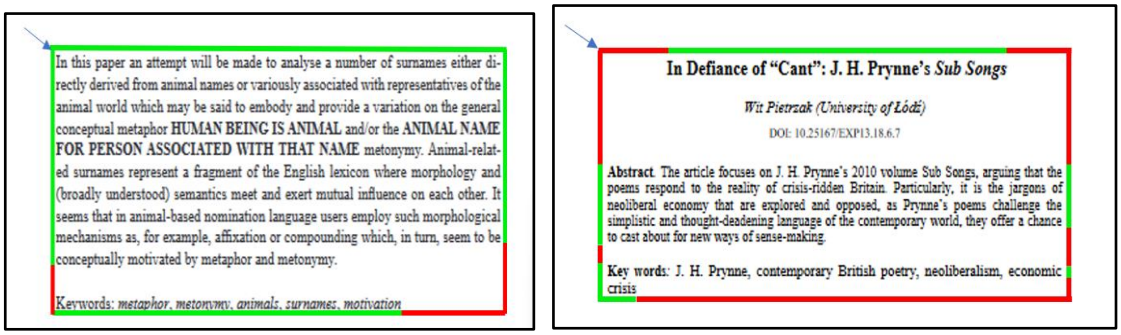


Figure 50. Independent and non-independent values

As the blue arrows indicate, when the upper left corner was present, i.e. the shape was made with green lines, a given token was perceived as independent. In contrast, when the upper left corner was absent, i.e. the shape was made with red lines, a given token was described as non-independent.

When it comes to completeness, a given token could be either complete or non-complete, depending on, in turn, the presence of the bottom right corner (see Figure 51).

COMPLETE NON-COMPLETE

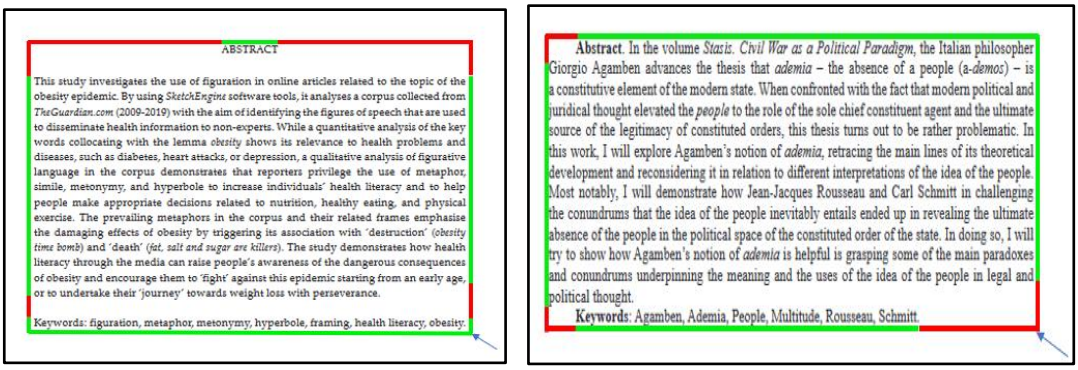
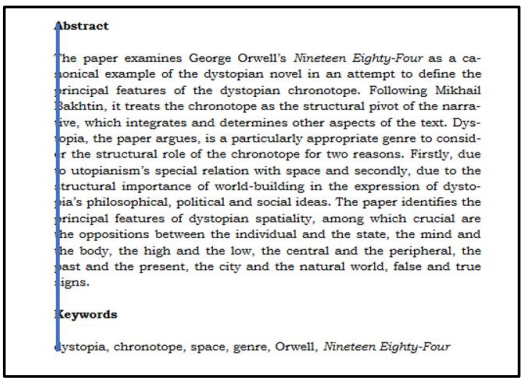


Figure 51. Complete and non-complete values

Likewise, if the bottom right corner was present, i.e. the shape was made with green lines, a given token was seen as complete whereas if the bottom right corner was absent, i.e. the shape was made with red lines, it was seen as non-complete.

Lastly, continuity was evaluated with reference to such values as continuous and non-continuous, which were established based on the alignment of all the parts, or the punctuation units, in a given token. In other words, if all the parts began in the same place, i.e. allowing to draw a straight line passing through their starting points, a given token was seen as continuous. If the parts began in different places within the page, a given token was perceived as non-continuous.

CONTINUOUS



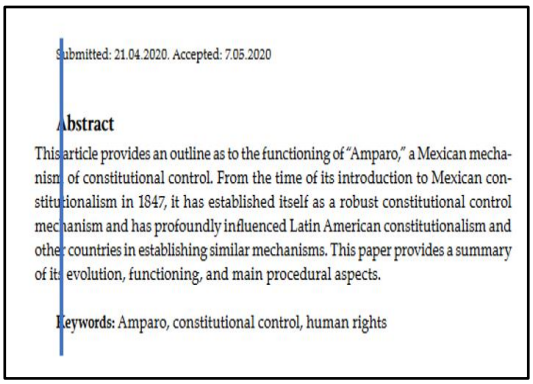
Abstract

The paper examines George Orwell's *Nineteen Eighty-Four* as a canonical example of the dystopian novel in an attempt to define the principal features of the dystopian chronotope. Following Mikhail Bakhtin, it treats the chronotope as the structural pivot of the narrative, which integrates and determines other aspects of the text. Dystopia, the paper argues, is a particularly appropriate genre to consider the structural role of the chronotope for two reasons. Firstly, due to utopianism's special relation with space and secondly, due to the structural importance of world-building in the expression of dystopia's philosophical, political and social ideas. The paper identifies the principal features of dystopian spatiality, among which crucial are the oppositions between the individual and the state, the mind and the body, the high and the low, the central and the peripheral, the past and the present, the city and the natural world, false and true signs.

Keywords

dystopia, chronotope, space, genre, Orwell, *Nineteen Eighty-Four*

NON-CONTINUOUS



Submitted: 21.04.2020. Accepted: 7.05.2020

Abstract

This article provides an outline as to the functioning of "Amparo," a Mexican mechanism of constitutional control. From the time of its introduction to Mexican constitutionalism in 1847, it has established itself as a robust constitutional control mechanism and has profoundly influenced Latin American constitutionalism and other countries in establishing similar mechanisms. This paper provides a summary of its evolution, functioning, and main procedural aspects.

Keywords: Amparo, constitutional control, human rights

Figure 52. Continuous and non-continuous values

At the same time, sequences of such constructions, analogous to linguistic constructions, can be represented in relation to three continua associated with the characterization of constructions. In other words, the corpus allows for observing that the sequence "independent, continuous, non-complete" is more fixed, i.e. it occurs more frequently, than the sequence "independent, continuous, complete". Additionally, after comparing the aforementioned sequences, it is possible to notice that they are different because "non-complete" is changed into "complete". Hence, based on such information, it is possible to establish more schematic constructions, e.g., "independent, continuous, completeness". Moreover, the corpus also enables consideration of the significance of punctuation units, based on the paralinguistic construction simplicity, thereby allowing for evaluating different degrees of complexity.

5.2. Concluding remarks

To sum up, Chapter 5 aimed to outline the foundational principles of the DISCOWER corpus, which was established with the intention of advancing research in two areas, i.e. construction-based approaches to academic ELF and a Cognitive (Construction) Grammar approach to discourse genres.

As it was mentioned in Section 5.1., selecting written ELF abstracts of research articles as corpus units enabled us to create a new (academic) ELF database which consists of a previously unavailable discourse genre construction. Simultaneously, embracing the framework of a Cognitive (Construction) Grammar was meant to facilitate further exploration within the evolving construction-based perspective on both written academic ELF and academic DGxC and take into account the potential of such form-

meaning pairings to combine with others, resulting in the emergence of more complex assemblies.

Considering such assemblies was, in turn, perceived as an attempt to integrate the assumptions of Linear Unit Grammar and Construction Grammar developed in academic ELF by incorporating Gestalt principles of grouping and the resulting establishment of groupings based on written data into academic ELF research. Concurrently, the description of these groupings was intended to facilitate their exploration not only by means of linguistic constructions but also, in particular, paralinguistic ones, i.e. constructions which hold great significance yet remain relatively unexplored in a Cognitive (Construction) Grammar approach to discourse genres.

Hence, although the corpus is still under development and some of its areas require improvements and/or further research, e.g. using eye-tracking methods to establish such groupings, the corpus constitutes a valuable source of data for, among others, further studies within the realm of academic ELF. In particular, further studies concerning the potential elaborated abstract construction based on linguistic and paralinguistic constructions seem valuable in order to determine its emerging nature, which could delineate a new academic discourse genre construction attributable to academic ELF disciplinary communities. These aspects are therefore the subject of further exploration presented in Chapter 6, where the main focus is on presenting the emerging elaborated construction with reference to both linguistic and paralinguistic constructions in three academic disciplines, i.e. law, linguistics, and literary studies.

Chapter 6

The emerging elaborated abstract construction in the academic ELF disciplinary communities of law, linguistics, and literary studies

6.0. Outline

The aim of Chapter 6 is to present the study conducted to identify and explore the emerging elaborated abstract construction (henceforth also EEAxC) in three academic ELF disciplinary communities, i.e. law, linguistics, and literary studies. The chapter focuses on describing the design of the study (Section 6.1.), providing a detailed discussion of the study itself (Section 6.2.), as well as presenting its findings (Section 6.3.) and identifying its limitations (Section 6.4.).

More precisely, Section 6.1. defines the research objectives guiding the study process and elucidates the rationale behind their formulation. At the same time, the section introduces the data utilized in the study and outlines the approach employed for their exploration.

Section 6.2. describes the study in a two-staged manner, i.e. presenting the identification and exploration of the EEAxC within the aforementioned disciplinary communities (henceforth also disciplines) and within more local and more international journal sub-communities in a particular discipline (henceforth also journals) (see Section 2.1.1. and 5.1.2.2.). At each stage, regularities concerning the construction under study are presented with reference to the three continua of features attributable to constructions in C(C)G (see Section 3.1.2.2.), i.e. the degree of fixedness, the degree of schematicity, and the degree of complexity.

Section 6.3. encapsulates the findings drawn from the two stages, discussing the similarities and differences among disciplinary communities and their sub-communities, while Section 6.4. identifies the constraints of the study to indicate areas that may require further refinement.

6.1. The design of the study

As indicated in Section 5.1., the compilation of the DISCOWER corpus was driven by the authors' aim to establish a database that could facilitate investigations contributing to

the advancement of research within construction-based approaches to academic ELF and the Cognitive (Construction) Grammar approach to (new) discourse genres (see Section 4.2.2.1. and 4.2.2.2.). Therefore, this study aims to leverage the above-mentioned potential of the DISCOWER corpus by identifying and exploring the EEAxC, i.e. a stabilizing discourse genre construction attributable to academic ELF disciplinary communities. However, before delving into a detailed discussion of the study process and its findings, its foundational assumptions should be presented. Hence, the next subsections discuss the design of the study, elucidating the research objectives (Section 6.1.1.), providing insights into the data used (Section 6.1.2.), and delineating the methodology adopted (Section 6.1.3.).

6.1.1. Research objectives

The present study is guided by four research objectives. More specifically, the study intends to:

1) identify the recurrent sequences of co-occurring linguistic and paralinguistic constructions signaling the emerging nature of the elaborated abstract construction;
2) explore the construction under study to delineate its regularities in terms of fixedness, schematicity, and complexity;
3) assess potential similarities and disparities in the above-mentioned regularities among three academic ELF disciplinary communities, i.e. law, linguistics, and literary studies;
4) assess potential similarities and disparities in the above-mentioned regularities among journal sub-communities in a given academic ELF disciplinary community, i.e. sub-communities disseminating research results in more local and more international journals.

The formulation of the first research objective is driven by two factors. Firstly, considering the appearance of new means of (ELF) research dissemination (see Section 2.2.2.1.2.) and the rather underexplored status of academic discourse genre constructions (see Section 4.3.), the study presented in this chapter is perceived as crucial for understanding the evolving practices in ELF scholarly communication and broadening the scope of studies on (new) discourse genre constructions. Secondly, given the fact that

knowledge pertaining to discourse genre constructions is seen (and thus explored) in relation to sequences of constructions of varying degrees of schematicity (see Section 3.1.2.2.2.), the establishment of the EEAxC is thus regarded as equivalent to identifying repetitive sequences encompassing more specific (linguistic) and more schematic (paralinguistic[126]) constructions constituting the above-mentioned assembly.

Next, building on the constructionist commitment of C(C)G (see Section 3.1.2.2.). as well as the studies on discourse genre constructions (see Section 4.2.), the second research objective results from the intention to provide a comprehensive account of the construction under investigation, i.e. by delineating its regularities in relation to the continua of features by means of which (discourse genre) constructions can be explored.

However, acknowledging the necessity to address potential diversity within (ELF) communities (see Section 2.2.2.2.2., Section 3.1.2.2.1., Section 4.2.2.1.), the study hinges on the assumption that the regularities of the EEAxC should be examined at (at least) two scales. More precisely, the formulation of the third research objective stems from the intention to account for the recognized heterogeneity among academic disciplinary ELF communities (see Section 2.2.2.2.2.), which, so far, has been mainly specified in relation to constructions within the sentence level[127] (see Section 2.3.).

Simultaneously, given the fact that generalizations about tendencies within a particular community should be established with caution, as the identified tendencies may be, in fact, characteristic of a sub-community (see Section 3.1.2.2.1), the formulation of the fourth research objective arises from the goal to determine the extent to which certain discipline-specific regularities result from an interplay between sub-communities within a particular discipline. To this aim, the study incorporates the distinction between sub-communities publishing in different types of journals, which is currently acknowledged

[126] As signaled in Section 5.1.4.2., since the importance of shapes in research on discourse genre constructions has been recognized, the study concentrates on the paralinguistic constructions pertaining to the shape of the EEAxC.

[127] Admittedly, prior to defining the research objectives and methodological assumptions presented in this study, several pilot studies were conducted, which included, among others, determining the relationship between the disciplines included in DISCOWER and the use of a particular shape-related construction. For instance, it was established that there are statistically significant differences between the frequency of occurrence of the complete construction in the disciplines under investigation. More specifically, it was established that the complete construction appears significantly more often in literary studies (23%) than in law (12%) and linguistics (7%) and there are statistically significant differences between law and linguistics (law/linguistics χ^2= 14,446, df = 1, $p < 0,05$; linguistics/literary studies; χ^2= 60,254, df = 1, $p < 0,05$; law/literary studies χ2= 22,564, df = 1, $p < 0,05$). Hence, such preliminary studies and their results also functioned as the motivation for further exploration of the differences among academic ELF disciplinary communities.

to be of great importance for further research on academic ELF communities (see Section 2.1.2.).

On this account, the study embraces the exploration of disciplinary communities and their sub-communities, which influences the extraction of data to be analyzed.

6.1.2. Data

As signaled above, the study relies on the data extracted from the DISCOWER corpus (see Chapter 5). Due to the research objectives discussed above, the data are divided into two groups, i.e. representing three academic disciplines (law, linguistics, and literary studies) and representing more local and more international journals in a particular discipline (e.g. local journals in law and international journals in law). In other words, the two groups in the study consist of three and six sub-groups, respectively.

Since the DISCOWER corpus includes information on the disciplinary affiliation of each token (see Section 5.1.4.), the total number of tokens representing the three disciplines can be determined based on the corpus webpage or a downloadable spreadsheet. Hence, the spreadsheet is downloaded and the total number of tokens in a given discipline is established. To identify a representative sample analyzed in this study, i.e. allowing for statistically valid generalizations, an online calculator is used (http://www.raosoft.com/samplesize.html)[128]. In accordance with the commonly employed practice[129], the value of the margin of error adopted in this study is 5% and the value of the confidence level is 95%. The information concerning the representative sample pertaining to each academic ELF disciplinary community is provided in Table 3 below.

DISCIPLINE	TOTAL NUMBER OF TOKENS IN THE CORPUS	SAMPLE SIZE
LAW	966	276
LINGUISTICS	817	262
LITERARY STUDIES	325	177
TOTAL	2108	715

[128] This calculator is used in, for instance, other construction-based studies dealing with the discourse level (see, for instance, Geka 2021).

[129] As noted by Truong (2024: 75) "the confidence level tells you how confident you are of this result. It is expressed as a percentage of times that different samples (if repeated samples were drawn) would produce this result. The common value for the confidence interval is 95%. A margin of error is a statistical measurement that accounts for the difference between actual and projected results. The common value for the margin of error is 5%".

Table 3. Representative samples

The information concerning the occurrence of tokens in more local or more international journals is, in turn, not available in the DISCOWER corpus. Hence, to identify such tokens, the data in the corpus are additionally tagged. To be more specific, based on the distinction suggested by Flowerdew (see Section 2.1.2.), more local journals are defined as those that contain not only articles in English but also articles in the native language of the country to which journals in DISCOWER are affiliated, i.e. Polish. Conversely, more international journals are defined as those that contain only articles in languages other than Polish (mainly English). Every issue of all the journals included in the corpus is checked in this respect, and the data are preserved in an additional column added to the spreadsheet, with the values "local" or "international" attributed to each token.

Given the aim to identify the extent to which the established regularities of a particular discipline result from the interplay of regularities pertaining to more local and international journals, separate representative samples are not established for the aforementioned sub-communities in a particular discipline. In other words, the data are first analyzed with respect to the three disciplines, and then the same data are used to analyze the disciplines from a narrower perspective, i.e. with respect to their sub-communities publishing in different types of journals.

Nevertheless, it is worth noting that the randomly sampled data reflect the general tendency of disciplines in the corpus, as differences between the entire population of local and international journals in a particular discipline in DISCOWER and the randomly sampled subset of tokens range from 1% to 4%. The details are presented in Table 4 below.

DISCIPLINE	**JOURNAL TYPE**	**TOTAL NUMBER OF TOKENS IN THE CORPUS**	**TOTAL NUMBER OF TOKENS IN THE SAMPLE**
LAW	LOCAL	810 (84%)	221 (80%)
	INTERNATIONAL	156 (16%)	55 (20%)
LINGUISTICS	LOCAL	127 (16%)	45 (17%)
	INTERNATIONAL	690 (84%)	217 (83%)
LITERARY STUDIES	LOCAL	219 (67%)	121 (68%)
	INTERNATIONAL	106 (33%)	56 (32%)

Table 4. Tokens representing local and international journals

The random sample derived from the corpus is established in several steps. To begin with, the modified spreadsheet is uploaded to Google Drive. Subsequently, the Google spreadsheet is filtered to extract the data representing law, linguistics, and literary studies by means of the values in the "discipline" column. Each obtained group is then sorted by clicking on "Data" and "Randomize range" to set a random list of tokens. Lastly, the first 276 rows in law, 262 rows in linguistics, and 177 rows in literary studies containing the data are copied and pasted into separate Google spreadsheets, constituting the dataset of three spreadsheets used for the first stage of the study. After this, each Google spreadsheet representing a particular discipline is filtered based on the values "local" and "international" in the added column. The data corresponding to these values are also copied and pasted into separate spreadsheets, which constitute the dataset comprised of six Google spreadsheets used in the second stage of the study.

To combine the data concerning linguistic and paralinguistic constructions available in the aforementioned Google spreadsheets, the CONCATENATE function is applied in a separate column. As a result, each token is described with reference to a sequence of constructions, e.g. abstract (label) || abstract content || keywords (label) || keywords (content), independent, continuous, non-complete, which are derived from the "composition" column, i.e. a column containing the data on the relationship between co-occurring linguistic constructions, and three separate columns pertaining to paralinguistic constructions, i.e. "independence", "continuity", and "completeness" (see Section 5.1.4.). Eventually, sequences established in this manner constitute the data by means of which the EEAxC is identified and explored.

6.1.3. Methodology

The study adopts a Cognitive (Construction) Grammar approach. Hence, building upon the cognitive (see Section 3.1.1.) and the constructionist commitment (see Section 3.1.2.) embraced in C(C)G, the EEAxC is perceived as a frame entailing (relations between) attributes and values coded by means of linguistic and paralinguistic forms (see Section 3.1.1.3.2.).

To represent the EEAxC as a form-meaning pairing, the study employs the method introduced by Langacker (see Section 3.1.2.2.3.), which, so far, has not been fully utilized in research on (academic) discourse genre constructions (see Section 4.3.). More specifically, the study posits that the representation of the EEAxC should reflect an interplay between linguistic and paralinguistic constructions and hence illustrate

groupings pertaining to graphological and semantic poles. In accordance with the principles of Cognitive (Construction) Grammar, the study assumes that the representation of the EEAxC can involve a more coarse-grained or a more fine-grained perspective.

To clarify, the study proposes that the EEAxC can be represented as either the superordinate punctuation unit constituting the EEAxC or the superordinate punctuation unit and its sub-units. An example of such a proposal is provided below, where the representation of the sequence "abstract (label) || abstract (content), independent, continuous, non-complete" is offered, illustrating a more coarse-grained (right side) and a more fine-grained (left side) perspective.

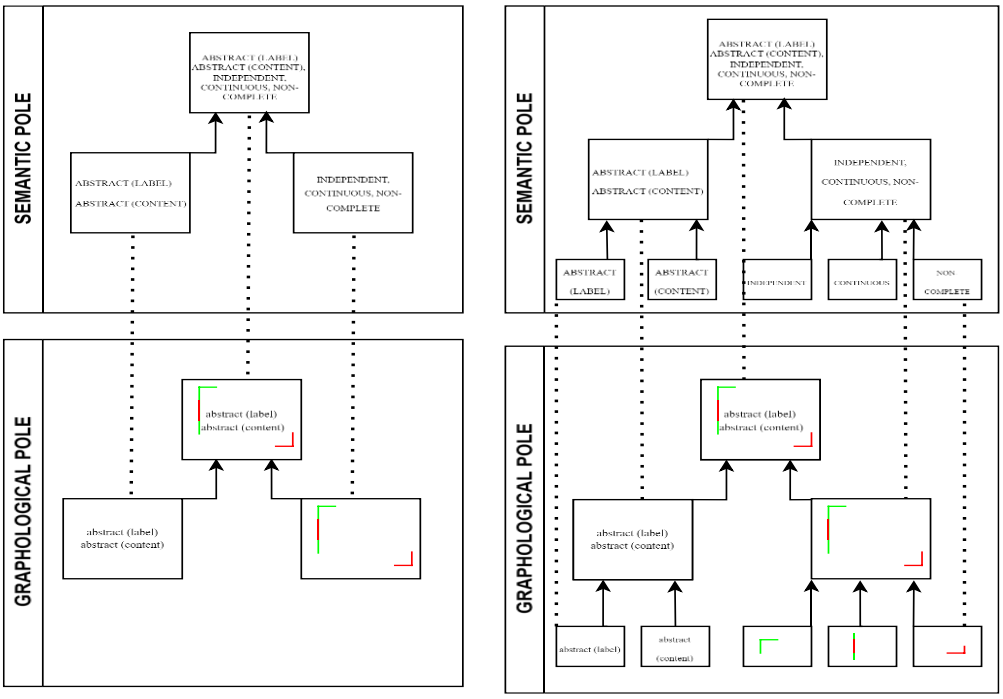


Figure 53. Ways of representing EAxC

A coarse-grained perspective depicts a single punctuation unit that includes abstract (label) and abstract (content) constructions, along with a representation of the paralinguistic constructions forming the shape in the adjacent box. In line with the solution adopted in C(C)G (see Section 3.1.2.1.), when a term, e.g. abstract (label), is used in reference to the semantic pole, uppercase letters are applied (ABSTRACT (LABEL)), whereas when a term is used to represent the graphological pole, uppercase letters are not used (abstract (label)). The representation of the paralinguistic constructions in the graphological pole entails the depiction of the upper left corner, left

edge and bottom right corner either in green (when the EEAxC encompasses the constructions independent, continuous, and/or complete) or red (when the EEAxC encompasses the constructions non-independent, non-continuous, and/or non-complete). The semantic pole, in turn, is represented by means of terms corresponding to their meaning, i.e. INDEPENDENT or NON-INDEPENDENT, CONTINUOUS or NON-CONTINUOUS, and COMPLETE, or NON-COMPLETE.

A finer-grained perspective additionally shows lower-level structures, indicating, e.g., that the abstract (label) and abstract (content) constructions occur in smaller punctuation units within a larger punctuation unit. Simultaneously, the overall representation of shape is divided into constructions corresponding to its particular fragments. Dotted lines are added in each case to indicate correspondences between graphological and semantic poles. However, for clarity of illustration, the finer-grained perspective presents correspondences between smaller punctuation units and paralinguistic constructions only in selected cases.

To identify and explore the EEAxC, the study relies on the clines by which constructions can be characterized[130], i.e. the degree of fixedness, the degree of schematicity, and the degree of complexity. In the first step, i.e. while describing the EEAxC with reference to the degree of fixedness, the main focus is put on identifying the EEAxC through the frequency with which sequences in a given community or sub-community appear[131]. Given the gradient view of constructionhood adopted in C(C)G, it is assumed that the degree of fixedness involves recognizing sequences occurring with varying frequencies, which may reflect possible differences in their stabilizing nature. Based on solutions adopted by scholars dealing with stabilities in academic ELF (see Section 2.2.2.2.2.), this study commences with the assumption that at least two occurrences of a given sequence indicate the EEAxC. As a consequence, the first step in each stage of the study involves distinguishing between emergent (occurring only once)

[130] The names of the clines are seen as equivalent to the ones introduced in Chapter 3, i.e. the fixedness-novelty continuum, the specificity-schematicity continuum, and the simplicity-complexity continuum. However, since the study concentrates on the emerging elaborated abstract construction, i.e. a stabilizing discourse genre construction, which is, among others, inherently complex, such names are employed to avoid the reference to, e.g., novelty and simplicity.

[131] Due to the fact that the study addresses the data representing academic ELF communities at various scales, the present exploration is thus seen as limited to identifying preferences appearing at the collective level, which are not considered indicative of an equivalent degree of entrenchment (see Section 3.1.2.2.1.).

and emerging (occurring at least twice[132]) sequences and determining which type is more common. However, considering the numerous sequences occurring twice, the second step is to identify the sequences comprising at least 5% of the total[133], which are subjected to a more detailed exploration in subsequent steps.

To clarify, a more detailed exploration entails the exploration of the identified sequences from (more) quantitative and (more) qualitative perspectives. When it comes to a more quantitatively-oriented approach, the analysis follows the practice adopted in studies on academic written ELF (see Section 2.2.2.2.2.) and discourse genre constructions (see Section 4.2.) and hence determines statistically significant differences between communities and sub-communities. For this purpose, two statistical tests are used: the chi-square test and Fisher's exact test. The chi-square test, which is considered popular for evaluating the relationship between nominal data, determines whether observed frequencies across variables significantly deviate from expected frequencies (Hirrel 2018). However, the application of the chi-square test is seen as inappropriate for small samples, e.g. when the expected frequencies are below 5 (Gorman and Johnson 2013, after Hirrel 2018). In such cases, Fisher's exact test is applied, which is "a popular choice for tables with expected values less than 5" (Cody 2019: 225). Consequently, the chi-square test is employed when expected values exceed 5, whereas Fisher's exact test is utilized when expected values are below 5. The identification of significant differences when comparing each considered pair is interpreted as an indication that a particular sequence is used much more frequently in a given community or sub-community, which may suggest that this emerging construction is specific to a given discipline or a journal within a discipline.

When it comes to a more qualitatively-oriented approach, this perspective is seen as related to the remaining continua, i.e. the degree of schematicity (the second step) and the degree of complexity (the third step). In the second step, an attempt is made to identify more schematic constructions relative to the ones recognized in the first step. For this purpose, the study adopts a roughly equivalent solution to the one embraced by scholars analyzing discourse genre constructions (see Section 4.2.2.1.) and concentrates on

[132] In the random sample data extracted from DISCOWER, there are no tokens representing the same author in the same journal. Consequently, each repetition of a given pattern indicates a degree of dispersion.
[133] In other words, this study adopts an approach roughly equivalent to the one taken by Hirrel (see Section 4.X). More precisely, Hirrel (2018: 94) suggests that describing all the data seems "impractical and uninteresting" and one of the ways in which Hirrel narrows down the scope of the presented data is by considering the frequency of occurrence. Therefore, the present study adopts the smallest threshold among the ones adopted by Hirrel, focusing on the data comprising at least 5% of the total.

delineating linguistic and/or paralinguistic constructions that appear in each recognized sequence. Concurrently, by applying such a method, the study establishes core attributes, i.e. the most frequently occurring attributes (see Figure 54).

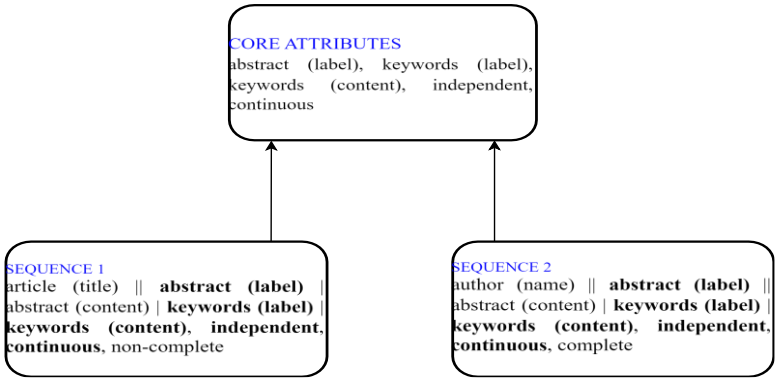


Figure 54. Identification of more schematic EAxCs

Next, a schematic representation of the EEAxC is created which reflects the core attributes identified in the second step.

In the third step, in turn, an attempt is made to describe the sequences identified in the first step as more or less complex. To achieve this, the study makes use of separators added to the composition, i.e. , i.e. "||" and "– –", to identify the boundaries of punctuation units. Subsequently, linguistic constructions occurring within the identified punctuation units are delineated, showing groupings pertaining to graphological and semantic poles (see Figure 55).

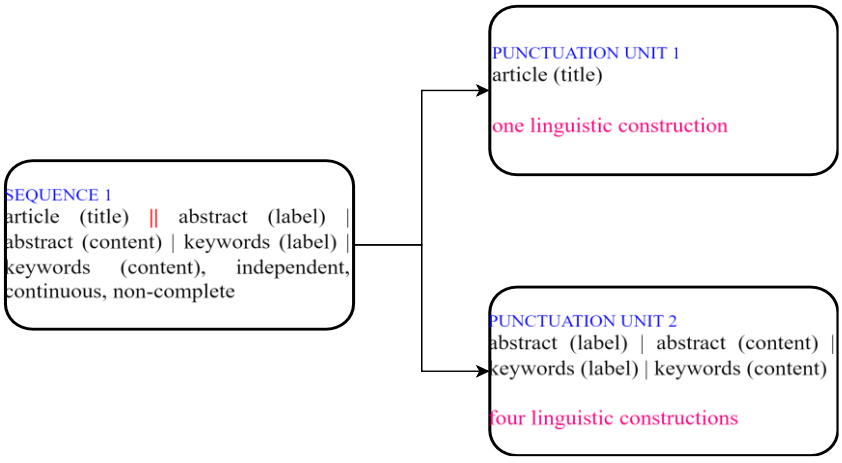


Figure 55. Identification of punctuation units

In this way, simultaneously, structural invariants, i.e. stable relations, between attributes are established[134]. To achieve this, the study focuses on regularities arising from the recurrent appearance of linguistic constructions either independently in a separate punctuation unit or together with different linguistic constructions within the same punctuation unit. For example, if keywords (label) and keywords (content) constructions always occur in the same punctuation unit in a given discipline or a journal, it is seen as an indication of the invariant relationship between them.

6.2. The study

In this section, the study process is discussed in detail. Since the study is divided into two stages, i.e. identification and exploration of the EEAxC in three academic ELF disciplinary communities (the first stage) and their sub-communities (the second stage), this section is divided into two corresponding subsections, i.e. the EEAxC in law, linguistics, and literary studies (Section 6.2.1.) and the EEAxC in local and international journals in law, linguistics, and literary studies (Section 6.2.2.).

6.2.1. The EEAxC in law, linguistics, and literary studies

The discussion concerning the identification and exploration of the EEAxC in the three disciplines is divided into three sub-sections, i.e. the degree of fixedness (Section 6.2.1.1.), the degree of schematicity (Section 6.2.1.2.), and the degree of complexity (Section 6.2.1.3.), to reflect the sequence of steps delineated in Section 6.1.3.

6.2.1.1. The degree of fixedness

To begin with, the first step of the study revealed that the disciplines under investigation can be characterized as having a tendency to produce emergent (rather than emerging) elaborated abstract constructions as emergent sequences outnumber emerging ones. More specifically, it was found that law can be characterized with reference to 101 sequences (with 64 emergent patterns). Likewise, linguistics can be described in relation to 68 sequences (with 43 emergent patterns) and literary studies in relation to 48 sequences (with 28 emergent patterns). The results are summarized in Table 5 below.

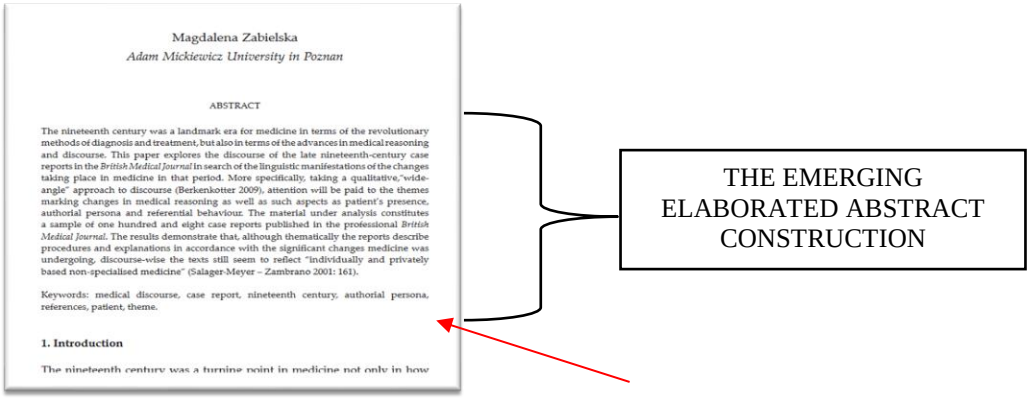
[134] Paralinguistic constructions are not taken into account in this step as each sequence is always characterized in relation to three paralinguistic constructions and their potentially stable relationships are already noticeable in the second step. To clarify, if, for example, the independent construction occurs with the continuous construction, they are defined as core attributes in the second step.

DISCIPLINE	EMERGENT SEQUENCES	% OF EMERGENT SEQUENCES	EMERGING SEQUENCES	% OF EMERGING SEQUENCES
LAW	64	63%	37	37%
LINGUISTICS	43	63%	25	37%
LITERARY STUDIES	28	58%	20	42%

Table 5. Emergent and emerging sequences in disciplines

Hence, a decision was made to perform a statistical analysis to determine whether there is a relationship between emergent sequences and a particular discipline. In other words, an attempt was made to determine whether the dominance of emergent sequences can be considered specific to a particular discipline or rather universal. The analysis indicated that this preference can be considered universal across disciplines under study, as no statistically significant differences between the disciplines were identified (law/linguistics χ^2= 0,000, df = 1, p > 0,05; linguistics/literary studies; χ^2= 0,285, df = 1, p > 0,05; law/literary studies χ^2= 0,349, df = 1, p > 0,05).

Concurrently, a closer examination of the emergent sequences indicated that the appearance of such sequences in the data was quite regular since emergent sequences frequently resulted from the presence of one linguistic or paralinguistic construction leading to the extension, i.e. deviation from the stabilizing unit. For example, when a recurrent sequence, such as “abstract (label) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete", was established, an emergent sequence frequently arose due to, for instance, the appearance of the complete construction in the above-mentioned sequence (see Figure 56).



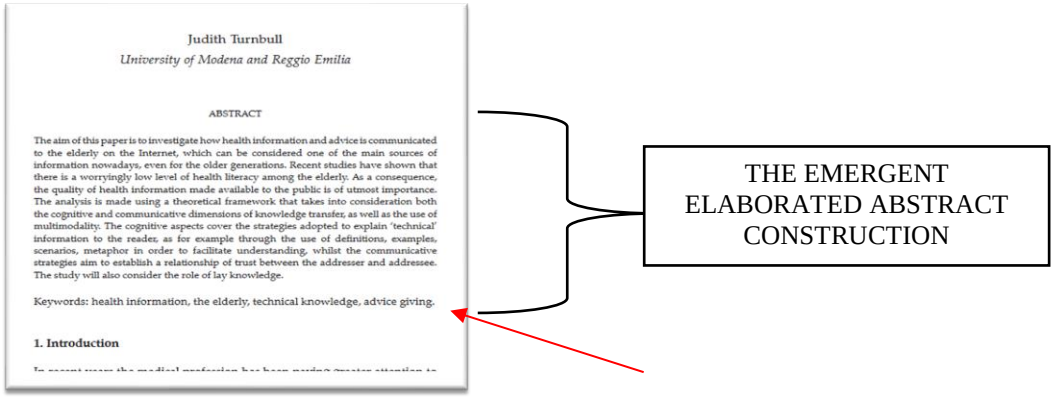


Figure 56. Emergent and emerging elaborated abstract constructions

When it comes to the focus of the present dissertation, i.e. the EEAxC, it was established that only 4 recurring sequences in law, 5 in linguistics and 3 in literary studies constitute at least 5% of the total. The emerging sequences are presented in Table 6 below.

	SEQUENCE	FREQUENCY	% OF THE DATA
LAW	1. abstract (label) – abstract (content) | keywords (label) – keywords (content), independent, continuous, non-complete	26	9%
	2. abstract (label) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete	25	9%
	3. article (title) || abstract (label: summary) || abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	17	6%
	4. author (name) | author (ORCID) (label) | author (affiliation) | author (contact) (content) || article (title) || abstract (label) – – abstract (content) || keywords (label) – keywords (content), independent, non-continuous, non-complete	13	5%
LINGUISTICS	5. abstract (label) | abstract (content) | keywords (label) – keywords (content), independent, continuous, non-complete	39	15%
	6. article (DOI) (content) | time log (received) (label+content) – time log (accepted) (label+content) || abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	34	13%
	7. keywords (label) – keywords (content) || abstract (label) || abstract (content), independent, continuous, non-complete	20	8%
	8. abstract (label) || abstract (content) || keywords (label) || keywords (content), independent, continuous, non-complete	18	7%
	9. abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	17	6%
LITERARY STUDIES	10. abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, complete	32	18%
	11. abstract (label) || abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	30	17%
	12. abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	12	7%
	13. abstract (label) | abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	9	5%

Table 6. Disciplines: sequences

Hence, the table indicates that the disciplines vary in terms of the degree of fixedness attributed to the emerging patterns as no homogeneous percentage data were established.

At the same time, the comparison of their frequencies across disciplines predominantly revealed statistically significant differences among the disciplinary communities, thereby indicating a preference for rather discipline-specific emerging sequences signaling the EEAxC. The details of the analysis are provided in Table 7 below, in which each number provided in the sequence column corresponds to the sequence numbered identically in Table 6 above.

DISCIPLINE	SEQUENCE	FREQ. IN LAW	FREQ. IN LIN	FREQ. IN LIT	LAW/LIN	LIN/LIT	LAW/LIT
LAW	**1**	26	0	0	$\chi^2 = 25{,}934$, df = 1, $p < 0{,}05$		$\chi^2 = 17{,}689$, df = 1, $p < 0{,}05$
	2	25	9	0	$\chi^2 = 7{,}178$, df = 1, $p < 0{,}05$	$p < 0{,}05$ (using Fisher's exact test)	$\chi^2 = 16{,}969$, df = 1, $p < 0{,}05$
	3	17	0	0	$\chi^2 = 16{,}664$, df = 1, $p < 0{,}05$	-	$\chi^2 = 11{,}327$, df = 1, $p < 0{,}05$
	4	13	0	0	$\chi^2 = 12{,}646$, df = 1, $p < 0{,}05$	-	$\chi^2 = 8{,}583$, df = 1, $p < 0{,}05$
LINGUISTICS (LIN)	**5**	0	39	0	$\chi^2 = 44{,}295$, df = 1, $p < 0{,}05$	$\chi^2 = 28{,}916$, df = 1, $p < 0{,}05$	-
	6	0	34	0	$\chi^2 = 38{,}233$, df = 1, $p < 0{,}05$	$\chi^2 = 25{,}898$, df = 1, $p < 0{,}05$	-
	7	0	20	0	$\chi^2 = 21{,}882$, df = 1, $p < 0{,}05$	$\chi^2 = 14{,}156$, df = 1, $p < 0{,}05$	-
	8	7	18	0	$\chi^2 = 5{,}698$, df = 1, $p < 0{,}05$	$\chi^2 = 12{,}680$, df = 1, $p < 0{,}05$	$p < 0{,}05$ (using Fisher's exact test)
	9	0	17	3	$\chi^2 = 18{,}493$, df = 1, $p < 0{,}05$	$\chi^2 = 5{,}582$, df = 1, $p < 0{,}05$	$p > 0{,}05$ (using Fisher's exact test)
LITERARY STUDIES (LIT)	**10**	0	0	32	-	$\chi^2 = 51{,}091$, df = 1, $p < 0{,}05$	$\chi^2 = 53{,}691$, df = 1, $p < 0{,}05$
	11	0	10	30	$p < 0{,}05$ (using Fisher's exact test)	$\chi^2 = 21{,}998$, df = 1, $p < 0{,}05$	$\chi^2 = 50{,}097$, df = 1, $p < 0{,}05$
	12	11	9	12	$\chi^2 = 0{,}114$, df = 1, $p > 0{,}05$	$\chi^2 = 2{,}594$, df = 1, $p > 0{,}05$	$\chi^2 = 1{,}747$, df = 1, $p > 0{,}05$
	13	0	9	9	$p < 0{,}05$ (using Fisher's exact test)	$\chi^2 = 0{,}731$, df = 1, $p > 0{,}05$	$p < 0{,}05$ (using Fisher's exact test)

Table 7. Disciplines: statistical analysis

Nevertheless, it was concurrently established that there is one universal stabilizing sequence for the three disciplines, i.e. "abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete". In other words, it was established that the universal EEAxC serves the function of summarizing the article and providing information about its key terms and its shape indicates its independence in relation to other linguistic constructions on the PDF page, equal status of all its parts (i.e.

lack of dependencies), and the lack of reaching its virtual limit. The universal EEAxC can thus be depicted as follows[135]:

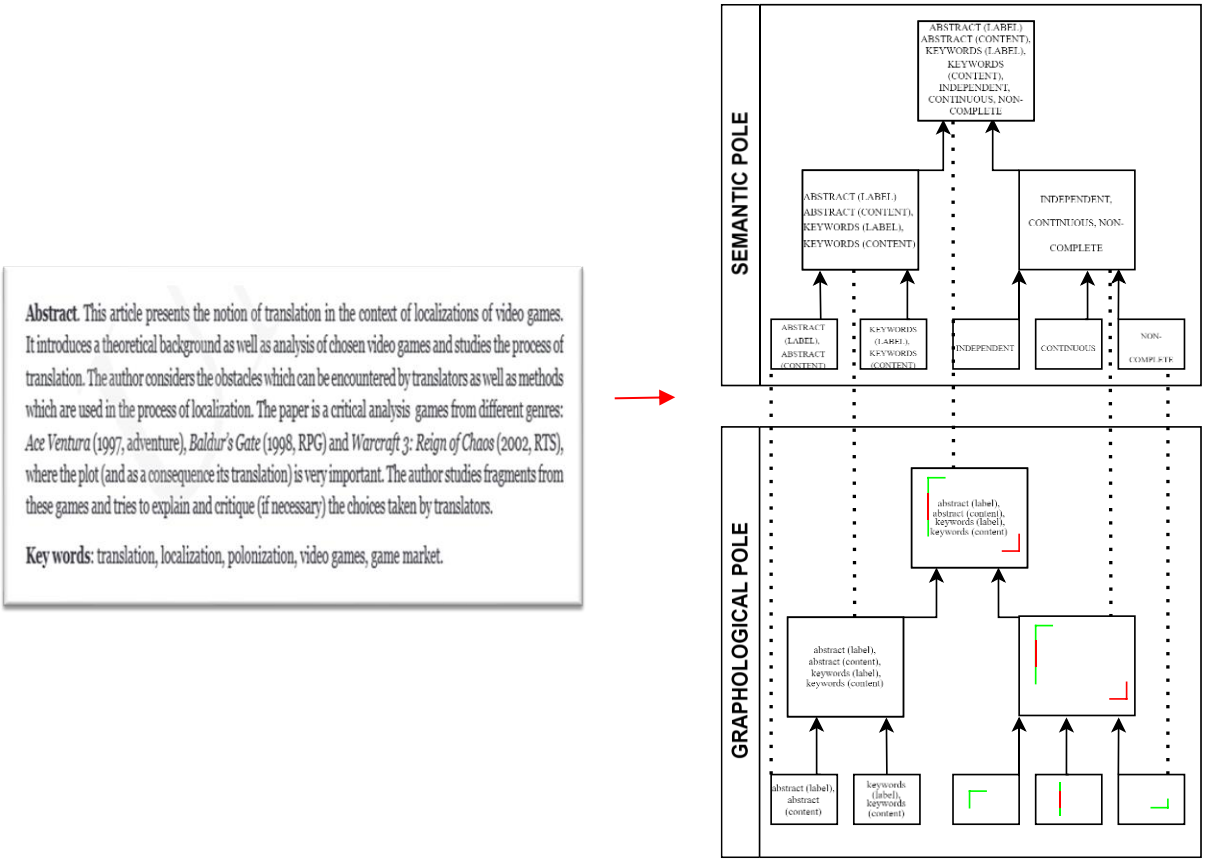


Figure 57: Disciplines: universal sequence

Additionally, it was established that there is one universal stabilizing sequence for linguistics and literary studies, i.e. "abstract (label) | abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete", and one for law and literary studies, i.e. "abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete", which are very similar in terms of their functions to the one of the EEAxC universal for the three disciplines. Their representations are provided below.

[135] To show what the representation stands for, the screenshot of a selected construct is provided. Henceforth, this solution is applied for all the fine-grained representations of the EEAxC to enhance clarity in the illustrations.

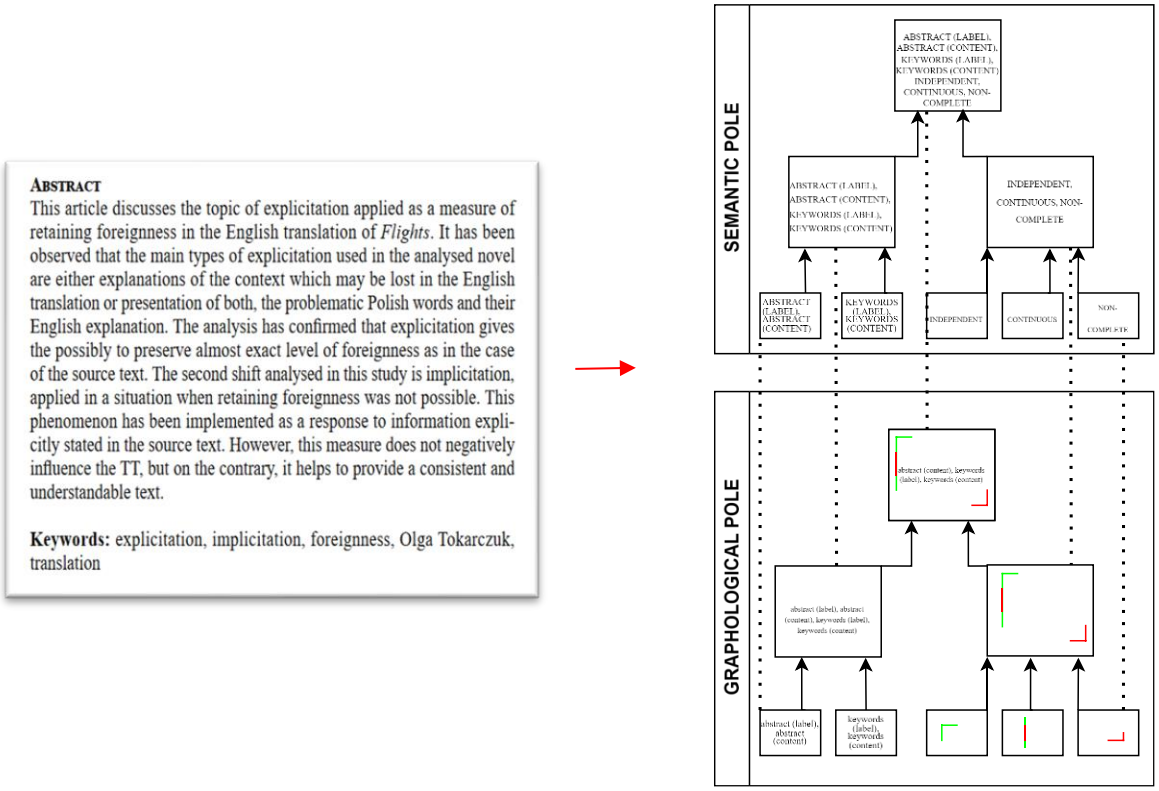


Figure 58. Universal sequence in linguistics and literary studies

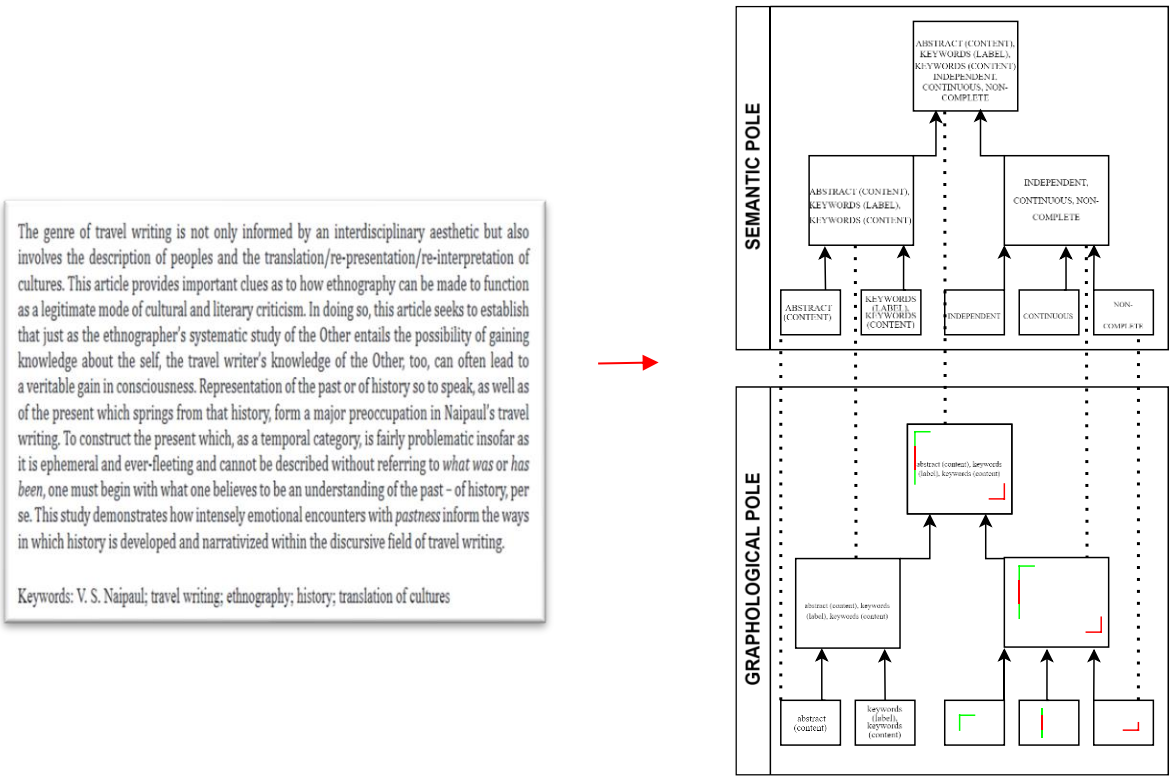


Figure 59. Universal sequence in law and literary studies

However, although the representation of the universal EEAxC for the three disciplines is identical to the universal representation of the EEAxC for linguistics and literary studies, it is worth noting that these sequences differ in the placement of the abstract (label) construction, which may affect its prominence (Langacker 2001). In other words, in the universal EEAxC in the three disciplines, the abstract (label) construction has a run-in position relative to the abstract (content) construction, whereas in the universal EEAxC in linguistics and literary studies, it is more separated, i.e. it is positioned above the abstract (content) construction. At the same time, it is worth noting that each established universal EEAxC can be characterized in relation to the same sequence of paralinguistic constructions, i.e. "independent, continuous, non-complete", suggesting the core attributes by means of which universal EEAxC can be described. A detailed discussion of core attributes within a given disciplinary community based on all the identified sequences is offered below.

6.2.1.2. The degree of schematicity

Following the procedure presented in Section 6.1.3, the second step of the exploration revealed that the disciplines under investigation differ in terms of more schematic EAxCs (and hence core attributes).

To begin with, it was found that law can be characterized with reference to 9 linguistic constructions and 5 paralinguistic constructions appearing in the identified sequences (see Table 8).

	CONSTRUCTIONS	**FREQUENCY PER ALL SEQUENCES**
LINGUISTIC	abstract (label)	3/4
	abstract (label: summary)	1/4
	article (title)	2/4
	author (affiliation)	1/4
	author (contact) (content)	1/4
	author (name)	1/4
	author (ORCID) (label)	1/4
	keywords (content)	4/4
	keywords (label)	4/4
PARALINGUISTIC	independent	3/4
	non-independent	1/4
	continuous	2/4
	non-continuous	2/4
	non-complete	4/4

Table 8. Linguistic and paralinguistic constructions in law

Hence, it was determined that the core attributes in law are: keywords (label), keywords (content), abstract (label_X), and non-complete. Based on these findings, it was

established that the more schematic EEAxC serves the function of summarizing the article and indicating its key terms, and its shape suggests the absence of reaching its virtual limit (see Figure 60).

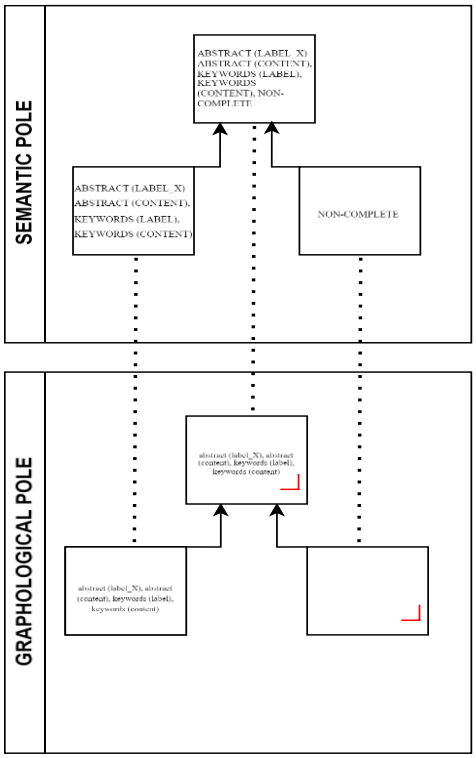


Figure 60. Schematic EEAxC in law

The use of _X with abstract (label) in the more schematic representation arises from the recognition that each sequence contains abstract (label) constructions. However, they differ in whether the label is “abstract” or “summary”, thus indicating the emergence of attribute-value relationships at a smaller scale, i.e. concerning labels. The same regularity can be determined with reference to keywords (label) and keywords (content) constructions, which function as the values of the emerging attribute keywords.

In linguistics, in turn, 6 linguistic constructions and only 3 paralinguistic constructions were identified (see Table 9).

	CONSTRUCTIONS	FREQUENCY PER ALL SEQUENCES
LINGUISTIC	abstract (label)	4/5
	article (DOI) (content)	1/5
	keywords (label)	5/5
	keywords (content)	5/5
	time log (accepted) (label+content)	1/5
	time log (received) (label+content)	1/5
PARALINGUISTIC	independent	5/5
	continuous	5/5
	non-complete	5/5

Table 9. Linguistic and paralinguistic constructions in linguistics

Therefore, it was established that the core attributes in linguistics are: keywords (label), keywords (content), independent, continuous, non-complete, leading to the emergence of a more schematic construction serving a function of summarizing the article and providing information about its key terms, and its shape indicates its independence in relation to other linguistic constructions on the PDF page, equal status of all its parts, and the lack of reaching its virtual limit. However, in this case, the abstract (content) construction is not obligatorily signaled by means of the abstract (label) construction (see Figure 61).

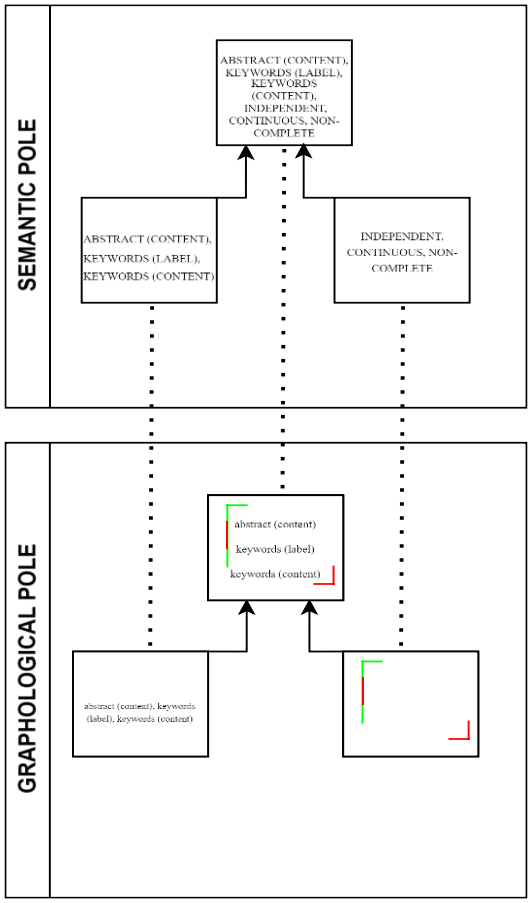


Figure 61. Schematic EEAxC in linguistics

In contrast to law, no attribute-value relationship was identified concerning the abstract (label) construction. However, a similar relationship was observed concerning keywords (label) and keywords (content) constructions.

Lastly, in literary studies, 3 linguistic and 4 paralinguistic constructions were determined (see Table 10).

	CONSTRUCTIONS	FREQUENCY PER ALL SEQUENCES
LINGUISTIC	abstract (label)	4/4
	keywords (label)	4/4
	keywords (content)	4/4
PARALINGUISTIC	independent	4/4
	continuous	4/4
	complete	1/4
	non-complete	3/4

Table 10. Linguistic and paralinguistic constructions in literary studies

Based on these data, the following core attributes were established: abstract (label), keywords (label), keywords (content), independent, continuous, leading to the emergence of a more schematic construction serving the function of summarizing the article and providing information about its key terms, and its shape indicates its independence from other linguistic constructions on the PDF page and equal status of all parts (see Figure 62).

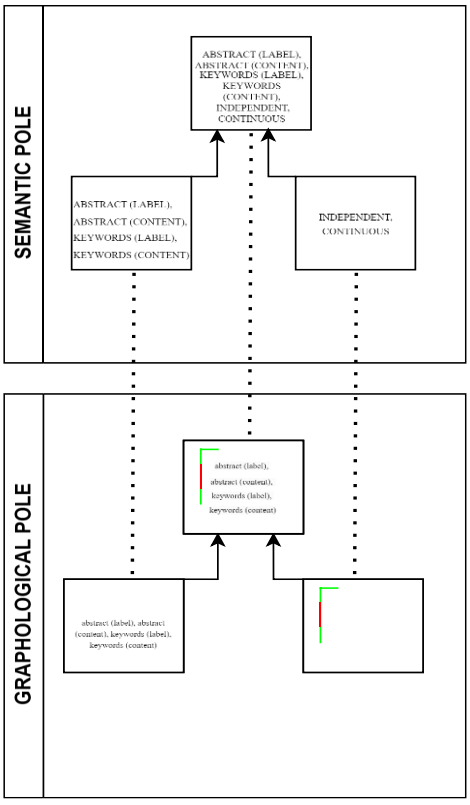


Figure 62. Schematic EEAxC in literary studies

Therefore, the same relationship was noted with reference to keywords (label) and keywords (content) constructions.

6.2.1.3. The degree of complexity

In this section, the results of the third step are presented, involving the division of sequences into punctuation units and the identification of structural invariants. To show the results of the exploration, a more fine-grained representation is offered in relation to the most frequent sequence in a given disciplinary community and the overall illustration of regularities is presented in tables showing the division into separate punctuation units.

When it comes to the regularities emerging in law, sequences ranged from 1 to 5 punctuation units, primarily encompassing various relationships between the abstract (label) and abstract (content) constructions (see Table 11).

SEQ.	P. UNIT 1	P. UNIT 2	P. UNIT 3	P. UNIT 4	P. UNIT 5
1	abstract (label) – abstract (content) | keywords (label) – keywords (content)				
2	abstract (label)	abstract (content)	keywords (label) – keywords (content)		
3	article (title)	abstract (label: summary)	abstract (content)	keywords (label) – keywords (content)	
4	author (name) | author (ORCID) (label) | author (affiliation) | author (contact) (content)	article (title)	abstract (label)	abstract (content)	keywords (label) – keywords (content)

Table 11. Punctuation units in law

However, based on Table 11, certain regularities were noticed. Firstly, the invariant relationship between keywords (label) and keywords (content) constructions was established as they always appear in the same punctuation. Secondly, it was noted that the article (title) construction always appears in a separate punctuation unit and does not link with other linguistic constructions. Out of all sequences, the most common in law is a sequence that can be considered the least complex, i.e. consisting of a single punctuation unit encompassing such constructions as abstract (label), abstract (content), keywords (label), and keywords (content) (see Figure 63).

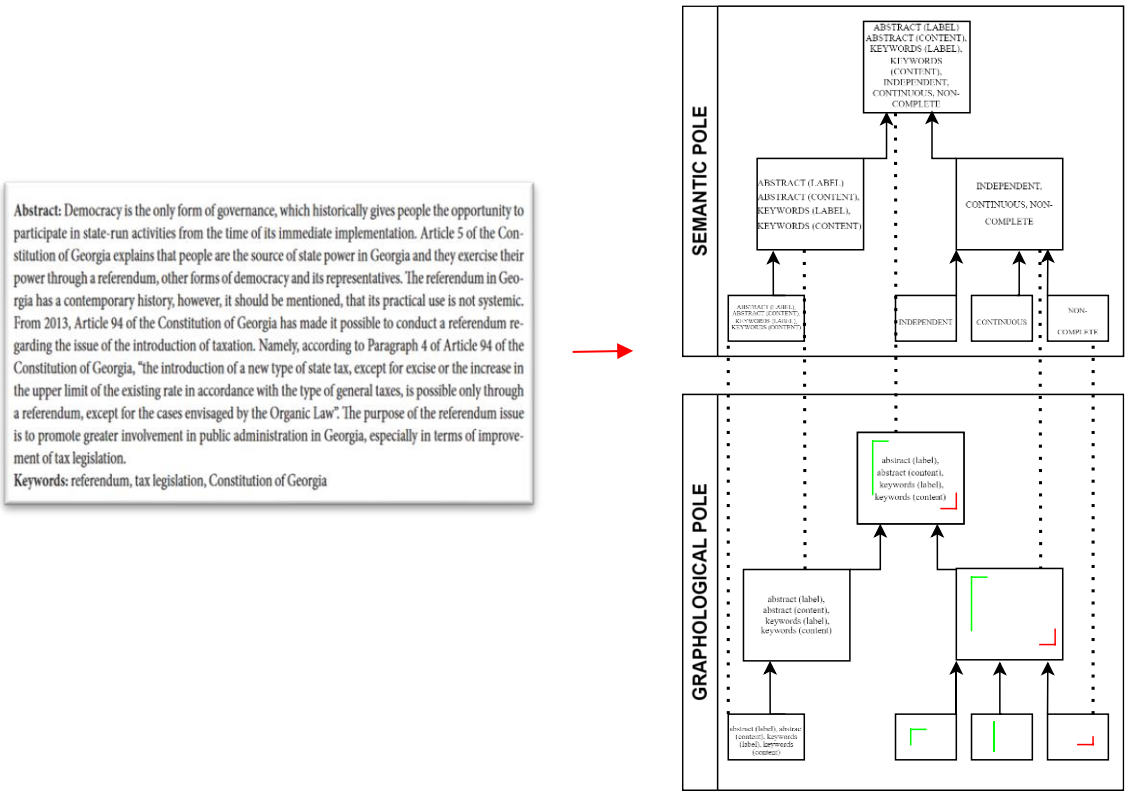


Figure 63. EEAxC in law

In linguistics, the sequences varied from 1 to 4 punctuation units, mirroring regularities seen in law, i.e. they primarily differ in the relationship between abstract (label) and abstract (content) constructions (see Table 12). However, unlike in law, no structural invariants were identified in this case.

SEQ.	P. UNIT 1	P. UNIT 2	P. UNIT 3	P. UNIT 4
5	abstract (label) | abstract (content) | keywords (label) – keywords (content)			
6	article (DOI) (content) | time log (received) (label+content) – time log (accepted) (label+content)	abstract (label) – abstract (content)	keywords (label) – keywords (content)	
7	keywords (label) – keywords (content)	abstract (label)	abstract (content)	
8	abstract (label)	abstract (content)	keywords (label)	keywords (content)
9	abstract (content)	keywords (label) – keywords (content)		

Table 12. Punctuation units in linguistics

At the same time, similar to law, the most frequent sequence was the simplest, i.e. comprising a single punctuation unit that includes constructions such as abstract (label), abstract (content), keywords (label), and keywords (content) (see Figure 64). However,

in linguistics, while the overall representation of this common construction is identical, a nuanced difference emerged regarding the positioning of the abstract (label) construction relative to the abstract (content) construction. Specifically, in this context, the abstract (label) construction is placed above the abstract (content) construction rather than in a run-in position (for a discussion see Section 6.2.1.1.).

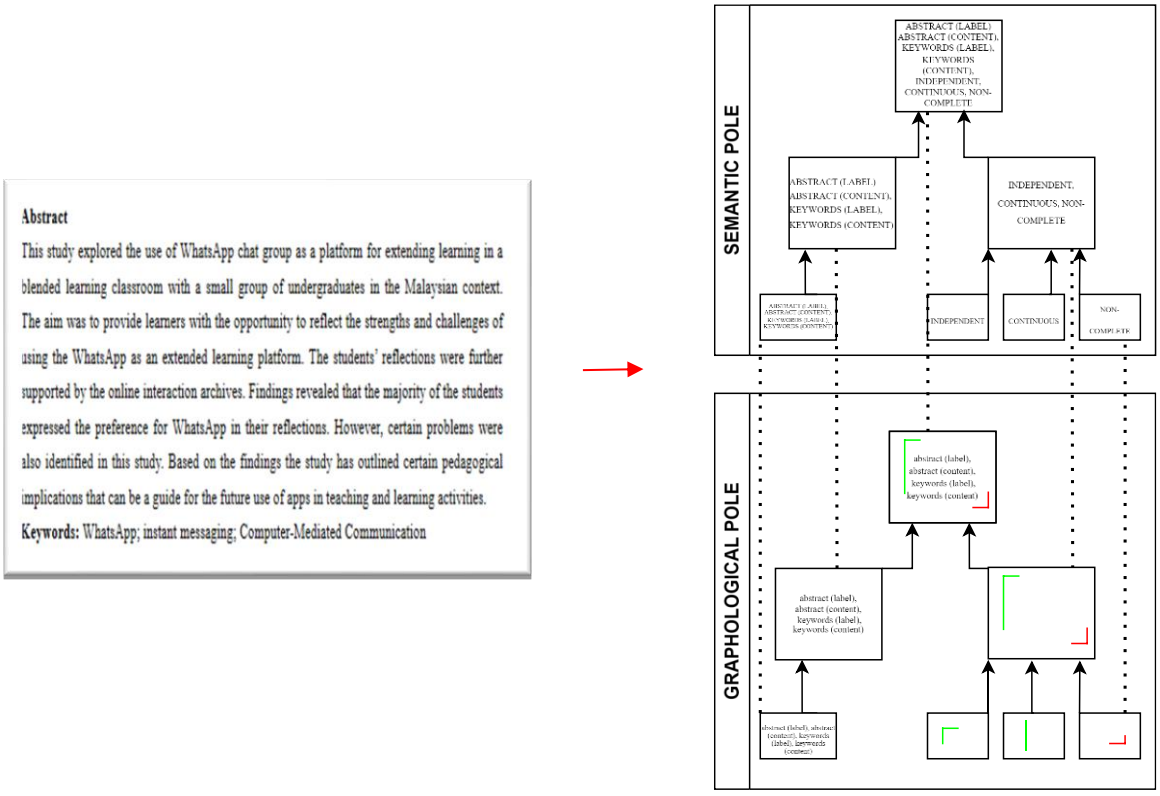


Figure 64. EEAxC in linguistics

In literary studies, in turn, the sequences ranged from 2 to 3 punctuation units, among which differences arose solely from the interplay between such constructions as abstract (label), abstract (content), keywords (label), and keywords (content) (see Table 13).

SEQ.	P. UNIT 1	P. UNIT 2	P. UNIT 3
10	abstract (label) – abstract (content)	keywords (label) – keywords (content)	
11	abstract (label)	abstract (content)	keywords (label) – keywords (content)
12	abstract (label) – abstract (content)	keywords (label) – keywords (content)	
13	abstract (label) | abstract (content)	keywords (label) – keywords (content)	

Table 13. Punctuation units in literary studies

In this case, it was established that each sequence exhibits an invariant relationship involving the co-occurrence of keywords (label) and keywords (content) constructions,

which always appear together in separate punctuation units. At the same time, unlike in law and linguistics, it was established that the most common sequence consists of two punctuation units (see Figure 65), encompassing two smaller groupings formed by such constructions as abstract (label), abstract (content), and keywords (label), keywords (content) respectively.

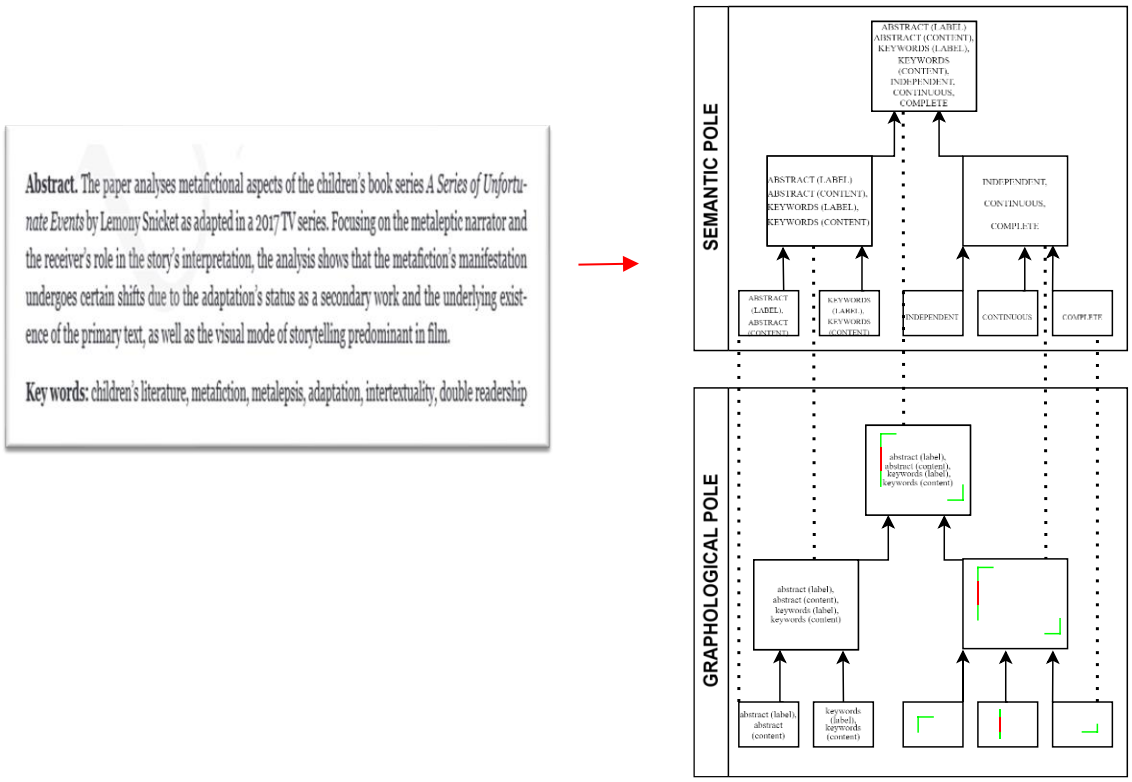


Figure 65. EEAxC in literary studies

6.2.2. The EEAxC in local and international journals in law, linguistics, and literary studies

In accordance with the division introduced in Section 6.2.1., the identification and exploration of the EEAxC in local and international journals in the three disciplines is organized into three sub-sections: the degree of fixedness (Section 6.2.2.1), the degree of schematicity (Section 6.2.2.2), and the degree of complexity (Section 6.2.2.3), following the steps outlined in Section 6.1.3. However, in this section, an attempt is made to assess the extent to which the regularities identified with reference to the communities publishing in more local and international journals reflect the regularities established during the exploration of disciplines.

6.2.2.1. The degree of fixedness

To begin with, the analysis revealed that the appearing tendency at the disciplinary level, indicating the dominance of emergent sequences, does not seem to be as homogeneous when viewed from a narrower perspective. More specifically, it was found that two types of journals in law can be similarly characterized, i.e. by a greater number of sequences deemed emergent (χ^2= 0,652, df = 1, p > 0,05). Conversely, in the case of linguistics, it was established that both local and international journals are characterized by a higher number of stabilizing sequences (χ^2= 1,329, df = 1, p > 0,05). In literary studies, in turn, it was shown that local journals exhibit more emergent sequences, while global journals display more emerging ones; however, significant differences between the sub-communities were not identified (χ^2= 1,947, df = 1, p > 0,05). The details are provided in Table 14 below.

	LAW		**LINGUISTICS**		**LITERARY STUDIES**	
	LOCAL	INTERNATIONAL	LOCAL	INTERNATIONAL	LOCAL	INTERNATIONAL
EMERGENT	50 (60%)	14 (70%)	12 (55%)	35 (69%)	16 (40%)	7 (60%)
EMERGING	33 (40%)	6 (30%)	10 (45%)	16 (31%)	24 (60%)	4 (36%)

Table 14. Emergent and emerging sequences in journals

Simultaneously, however, the factors influencing the appearance of emergent sequences can be considered the same as in the disciplines under investigation (see Section 6.2.1.1.).

Among the sequences constituting 5% of the total, both new and already discussed patterns (see Section 6.2.1.1.) were identified. In local law journals, 5 sequences were identified for further analysis, while in international journals 4 were noticed (see Table 15).

LAW	**SEQUENCE**	**FREQUENCY**	**% OF THE DATA**	**RECOGNITION AT THE DISCIPLINE LEVEL**
LOCAL	1. abstract (label) – abstract (content) | keywords (label) – keywords (content), independent, continuous, non-complete	26	12%	yes
	2. article (title) || abstract (label: summary) || abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	17	7,7%	yes
	3. author (name) | author (ORCID) (label) | author (affiliation) | author (contact) (content) || article (title) || abstract (label) – - abstract (content) || keywords (label) – keywords (content), independent, non-continuous, non-complete	13	6%	yes
	4. abstract (label: summary) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete	12	5%	no

	5. abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	11	5%	no
INTERNATIONAL	6. abstract (label) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete	18	34%	yes
	7. journal (title) | issue (number) || article (title) || author (name) || abstract (label) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete	9	16%	no
	8. abstract (label) || abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	7	13%	no
	9. journal (title) | issue (number) – issue (page range) | article (DOI) (label) – article (DOI) (content) || article (title) || author (name) || abstract (label) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete	3	5%	no

Table 15. Emerging sequences in local and international journals in law

In local journals, most of the data consisted of established disciplinary-level information. Conversely, global journals predominantly featured newly identified sequences. Additionally, statistical analysis indicated that three sequences can be considered universally applicable in the discipline:

1) author (name) | author (ORCID) (label) | author (affiliation) | author (contact) (content) || article (title) || abstract (label) – – abstract (content) || keywords (label) – keywords (content), independent, non-continuous, non-complete, i.e. the EEAxC summarizing the article and providing information about its key terms, its title, as well as the details about its author(s), and its shape indicates its independence in relation to other linguistic constructions on the PDF page, non-equal status of all its parts, and the lack of reaching its virtual limit;
2) abstract (label: summary) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete, i.e. the EEAxC summarizing the article and providing information about its key terms, and its shape indicates its dependence in relation to other linguistic constructions on the PDF page, non-equal status of all its parts, and the lack of reaching its virtual limit;
3) abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete, i.e. the EEAxC summarizing the article and providing information about its key terms, and its shape indicates its independence in relation to other linguistic constructions on the PDF page, equal status of all its parts, and the lack of reaching its virtual limit.

The representation of the most frequently occurring one is provided below.

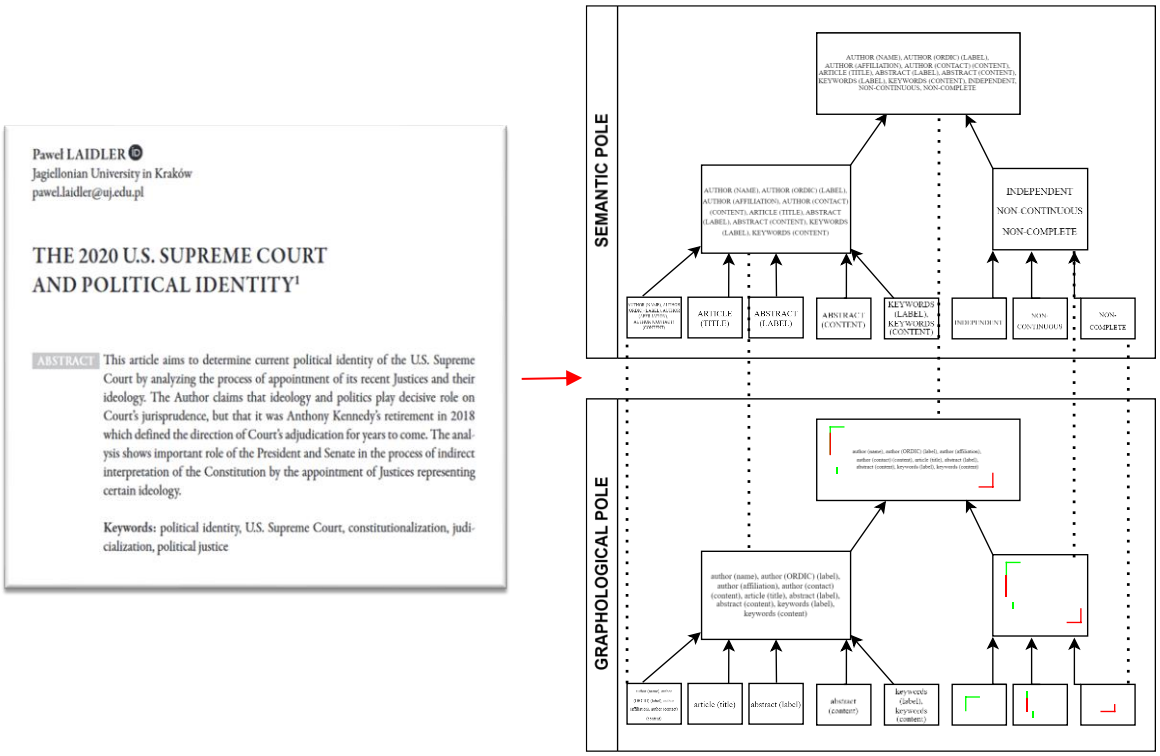


Figure 66. Universal EEAxCs in law

When it comes to the remaining sequences, it was established that they should be viewed as specific to a given sub-community (see Table 16).

LAW	SEQUENCE	FREQ. IN L	FREQ. IN I	L/I
LOCAL (L)	1	26	0	$\chi^2 = 7{,}144$, df = 1, $p < 0{,}05$
	2	17	0	$p < 0{,}05$ (using Fisher's exact test)
	3	13	0	$p > 0{,}05$ (using Fisher's exact test)
	4	12	0	$p > 0{,}05$ (using Fisher's exact test)
	5	11	0	$p > 0{,}05$ (using Fisher's exact test)
INTERNATIONAL (I)	6	7	18	$\chi^2 = 46{,}715$, df = 1, $p < 0{,}05$
	7	0	9	$p < 0{,}05$ (using Fisher's exact test)
	8	0	7	$p < 0{,}05$ (using Fisher's exact test)
	9	0	3	$p < 0{,}05$ (using Fisher's exact test)

Table 16. Differences between local and international journals in law

In linguistics, in turn, 8 sequences were identified in local journals and 6 sequences in global journals (see Table 17 below).

LINGUISTICS	SEQUENCE	FREQUENCY	% OF THE DATA	RECOGNITION AT THE DISCIPLINE LEVEL
LOCAL	1. journal (title) || journal (affiliation) | journal (ISSN) – issue (number) – issue (page range) | article (DOI) (content) | journal (website) || author (name) | author (affiliation) | author (ORCID) (content) | author (contact) (content) || article (title) || abstract (label) ||	5	11%	no

	abstract (content) || keywords (label) – keywords (content) || keywords (label) (Polish) – keywords (content) (Polish), non-independent, non-continuous, non-complete			
	2. abstract (label) || article (title) || abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	4	9%	no
	3. keywords (label) – keywords (content) || abstract (label) – abstract (content), non-independent, continuous, non-complete	3	7%	no
	4. article (title) || abstract (label: summary) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete	3	7%	no
	5. abstract (label) || abstract (content), independent, continuous, non-complete	3	7%	no
	6. abstract (label) || abstract (content), non-independent, continuous, non-complete	3	7%	no
	7. article (title) – – issue (page) || abstract (label) || abstract (content) || abstract (label: summary) (Polish) || abstract (content) (Polish), non-independent, non-continuous, non-complete	3	7%	no
	8. abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	3	7%	no
INTERNATIONAL	9. abstract (label) | abstract (content) | keywords (label) – keywords (content), independent, continuous, non-complete	37	17%	yes
	10. article (DOI) (content) | time log (received) (label+content) – time log (accepted) (label+content) || abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	34	16%	yes
	11. keywords (label) – keywords (content) || abstract (label) || abstract (content), independent, continuous, non-complete	20	9%	yes
	12. abstract (label) || abstract (content) || keywords (label) || keywords (content), independent, continuous, non-complete	18	8%	yes
	13. abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	17	8%	yes
	14. abstract (label) || abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	10	5%	no

Table 17. Emerging sequences in local and international journals in linguistics

As indicated in Table 17, all identified sequences in local journals were new compared to the ones identified in the first stage of the study. In international journals, however, only one new sequence was established. Simultaneously, statistical analysis indicated that three sequences can be considered universal, i.e.:

1) abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete;
2) abstract (label) || abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete;
3) abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete.

Hence, the three EEAxCs can be described as summarizing the article and providing information about its key terms, and their shapes indicate their independence in relation to other linguistic constructions on the PDF page, equal status of all its parts and the lack of reaching its virtual limit. The only difference emerges from the lack of the abstract (label) construction in the first sequence, which, in fact, is the one that occurs most frequently (see Figure 67).

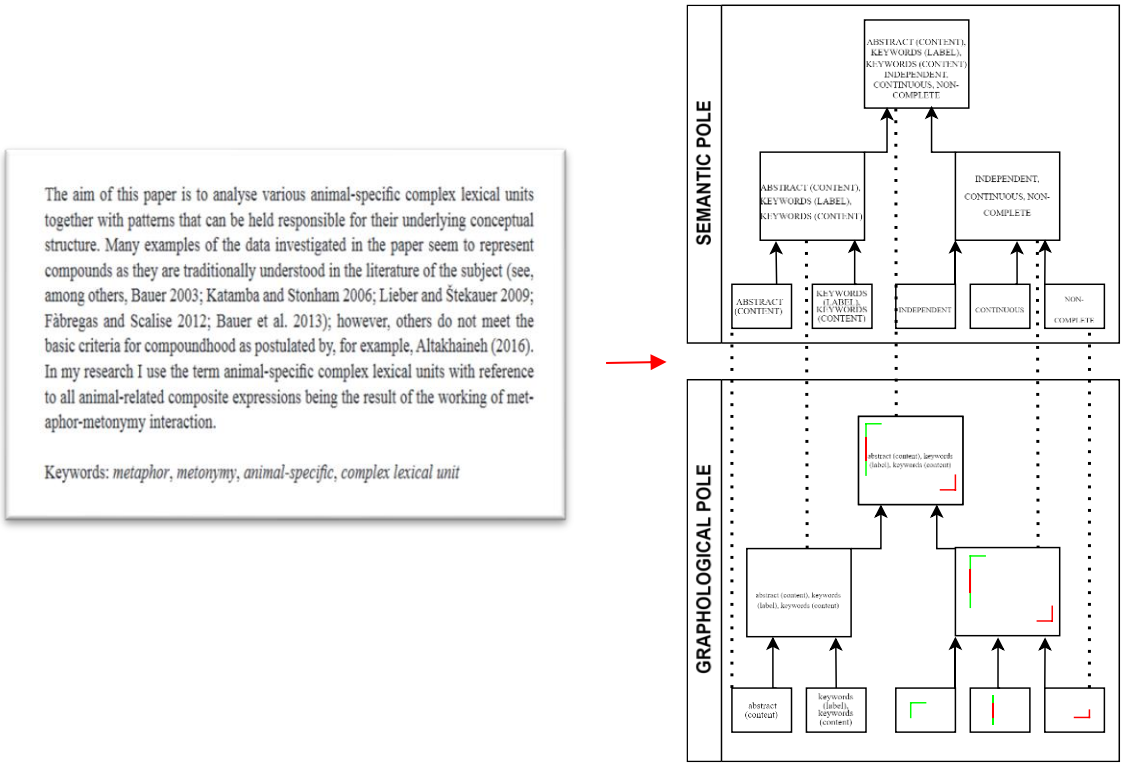


Figure 67. Universal EEAxCs in linguistics

For the remaining sequences, it was determined that they should be considered specific to a particular sub-community (see Table 18).

LINGUISTICS	SEQUENCE	FREQ. IN L	FREQ. IN I	L/I
LOCAL (L)	1	5	0	$p < 0,05$ (using Fisher's exact test)
	2	4	0	$p < 0,05$ (using Fisher's exact test)
	3	3	0	$p < 0,05$ (using Fisher's exact test)
	4	3	0	$p < 0,05$ (using Fisher's exact test)
	5	3	0	$p < 0,05$ (using Fisher's exact test)
	6	3	0	$p < 0,05$ (using Fisher's exact test)
	7	3	0	$p < 0,05$ (using Fisher's exact test)
	8	3	6	$p > 0,05$ (using Fisher's exact test)
INTERNATIONAL (I)	9	0	37	$\chi^2 = 8,935$, df = 1, $p < 0,05$
	10	0	34	$\chi^2 = 8,102$, df = 1, $p < 0,05$
	11	0	20	$p < 0,05$ (using Fisher's exact test)
	12	0	18	$p < 0,05$ (using Fisher's exact test)
	13	0	17	$p > 0,05$ (using Fisher's exact test)
	14	0	10	$p > 0,05$ (using Fisher's exact test)

Table 18. Differences between local and international journals in linguistics

In literary studies, as in the case of linguistics, 8 sequences were recognized in local journals and 6 in international journals (see Table 19 below).

LITERARY STUDIES	SEQUENCE	FREQUENCY	% OF THE DATA	RECOGNITION AT THE DISCIPLINE LEVEL
LOCAL	1. abstract (label) ‖ abstract (content) ‖ keywords (label) – keywords (content), independent, continuous, non-complete	25	21%	yes
	2. abstract (label) | abstract (content) ‖ keywords (label) – keywords (content), independent, continuous, non-complete	9	7%	yes
	3. article (title) ‖ author (name) | author (affiliation) ‖ abstract (label) | abstract (content) ‖ keywords (label) – keywords (content), independent, non-continuous, non-complete	8	7%	no
	4. abstract (label: summary) ‖ abstract (content), independent, non-continuous, non-complete	8	7%	no
	5. article (title) ‖ article (DOI) (label) – article (DOI) (content) ‖ author (name) – author (affiliation) | author (ORCID) (label) – author (ORCID) (content) ‖ abstract (label) – abstract (content) ‖ keywords (label) – keywords (content), non-independent, non-continuous, non-complete	8	7%	no
	6. abstract (label) – abstract (content) ‖ keywords (label) – keywords (content), independent, continuous, non-complete	6	5%	yes
	7. abstract (label) | abstract (content), independent, continuous, non-complete	6	5%	no
	8. article (title) ‖ abstract (label: summary) ‖ abstract (content) ‖ keywords (label) – keywords (content) ‖ keywords (label) (Polish) – keywords (content) (Polish), non-independent, non-continuous, non-complete	6	5%	no
INTERNATIONAL	9. abstract (label) – abstract (content) ‖ keywords (label) – keywords (content), independent, continuous, complete	30	54%	yes
	10. abstract (label) – abstract (content) ‖ keywords (label) – keywords (content), independent, continuous, non-complete	6	11%	yes
	11. abstract (label) ‖ abstract (content) ‖ keywords (label) – keywords (content), independent, continuous, non-complete	5	9%	yes
	12. abstract (content) ‖ keywords (label) – keywords (content), independent, continuous, non-complete	3	5%	no
	13. author (name) ‖ author (affiliation) – author (contact) (content) | author (ORCID) (content) ‖ article (title) ‖ abstract (label) – abstract (content) ‖ keywords (label) – keywords (content), non-independent, non-continuous, non-complete	3	5%	no
	14. author (name) ‖ author (affiliation) – author (contact) (content) ‖ article (title) ‖ abstract (label) – abstract (content) ‖ keywords (label) – keywords (content), non-independent, non-continuous, non-complete	3	5%	no

Table 19. Emerging sequences in local and international journals in literary studies

When it comes to the sequences in local journals, the majority of them constituted newly recognized sequences. In international journals, however, there were 3 sequences recognized in the first stage of the study and 3 new ones. A statistical analysis indicated that all the sequences identified in the local journals as well as 2 recognized in the international journals (Sequence 10 and Sequence 11) can be perceived as universal.

Nevertheless, the representation of the most frequently occurring EEAxC which can be considered universal is provided below.

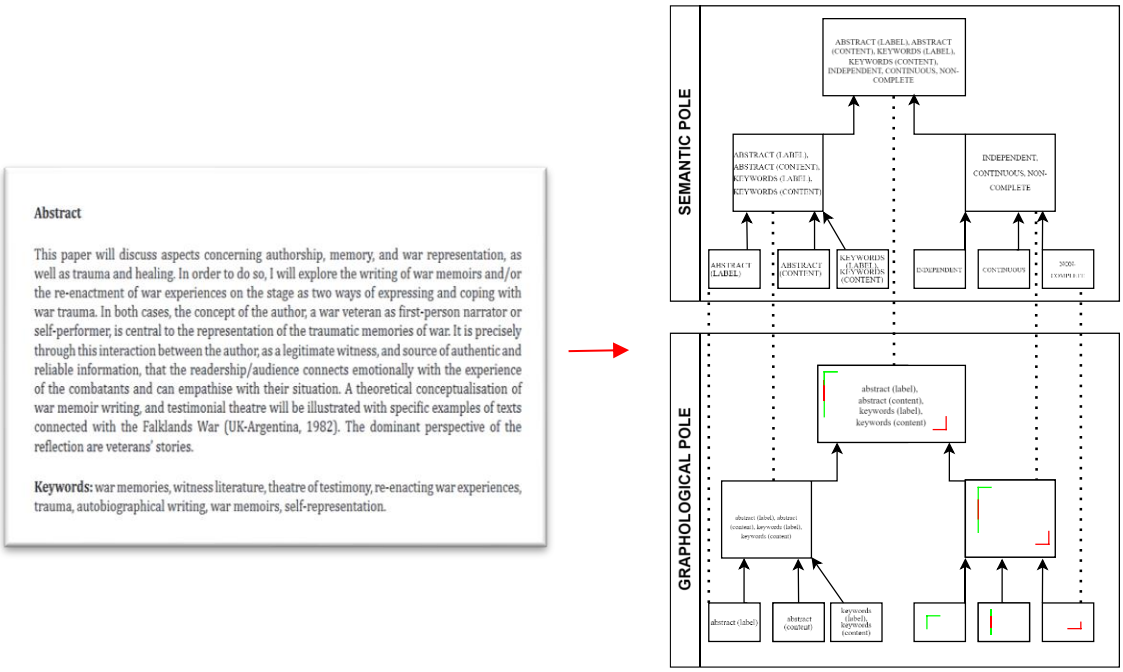


Figure 68. Universal EEAxCs in literary studies

The details of the statistical analysis are provided in Table 20.

LITERARY STUDIES	SEQUENCE	FREQ. IN L	FREQ. IN I	L/I
LOCAL (L)	1[136]	25	5	$\chi^2 = 3{,}744$, df = 1, $p > 0{,}05$
	2	9	0	$p > 0{,}05$ (using Fisher's exact test)
	3	8		
	4	8	0	$p > 0{,}05$ (using Fisher's exact test)
	5	8	0	$p > 0{,}05$ (using Fisher's exact test)
	6	6	0	$p > 0{,}05$ (using Fisher's exact test)
	7	6	0	$p > 0{,}05$ (using Fisher's exact test)
	8	6	0	$p > 0{,}05$ (using Fisher's exact test)
INTERNATIONAL (I)	9	2	30	$\chi^2 = 69{,}674$, df = 1, $p < 0{,}05$
	10	6	6	$p > 0{,}05$ (using Fisher's exact test)
	11	25	5	$\chi^2 = 3{,}744$, df = 1, $p > 0{,}05$
	12	0	3	$p < 0{,}05$ (using Fisher's exact test)
	13	0	3	$p < 0{,}05$ (using Fisher's exact test)
	14	0	3	$p < 0{,}05$ (using Fisher's exact test)

Table 20. Differences between local and international journals in literary studies

[136] The same sequence, i.e. 1 and 11, was established in local and international journals. However, for the clarity of description, the results of statistical analysis are repeated to achieve the same order of sequences in Table 20 and in Table 22.

Overall, the narrower perspective enabled the discovery of a few new sequences that can be universally attributed to the disciplines under investigation or that can be seen as specific to their sub-communities. Nevertheless, it seems that the universal sequences defined in linguistics are the most consistent with the universal sequences established in the first stage of the study.

6.2.2.2. The degree of schematicity

When it comes to the degree of schematicity, the exploration showed that the schematic constructions in law, linguistics, and literary studies identified in the first stage can be attributed to local journals in law and international journals in linguistics. When it comes to international journals in law, local journals in linguistics and both local and international journals in literary studies, different schematic constructions were delineated.

To begin with, it was discovered that the local journals in law can be characterized by means of 9 linguistic and 5 paralinguistic constructions. Simultaneously, 10 linguistic and 5 paralinguistic constructions were identified in the international journalsin law (see Table 21). However, as the regularities of the local journals mirror the ones described for the discipline (see Section 6.2.1.2.), this section focuses on the regularities identified in relation to the international ones.

	CONSTRUCTIONS (LOCAL)	**FREQ. PER ALL SEQUENCES (LOCAL)**	**CONSTRUCTIONS (INTERNATIONAL)**	**FREQ. PER ALL SEQUENCES (INTERNATIONAL)**
LINGUISTIC	abstract (label)	3/5	abstract (label)	4/4
	abstract (label: summary)	2/5	article (DOI) (content)	1/4
	article (title)	2/5	article (DOI) (label)	1/4
	author (affiliation)	1/5	article (title)	2/4
	author (contact) (content)	1/5	author (name)	2/4
	author (name)	1/5	issue (number)	2/4
	author (ORCID) (label)	1/5	issue (page range)	1/4
	keywords (content)	5/5	journal (title)	2/4
	keywords (label)	5/5	keywords (content)	4/4
			keywords (label)	4/4
PARALINGUISTIC	independent	4/5	independent	1/4
	non-independent	1/5	non-independent	3/4
	continuous	3/5	continuous	1/4
	non-continuous	2/5	non-continuous	3/4
	non-complete	5/5	non-complete	4/4

Table 21. Linguistic and paralinguistic constructions in local and international journals in law

Based on the table above, the following core attributes were identified in relation to the international journals in law: keywords (label), keywords (content), abstract (label), and non-complete. Therefore, the only difference concerning these regularities and the regularities of the local journals and the discipline as a whole is that each sequence in the international journals is characterized by the presence of the abstract (label) construction, which is not interchangeably used with the abstract (label: summary) construction. On this account, roughly the same schematic representation can be determined (see Figure 69).

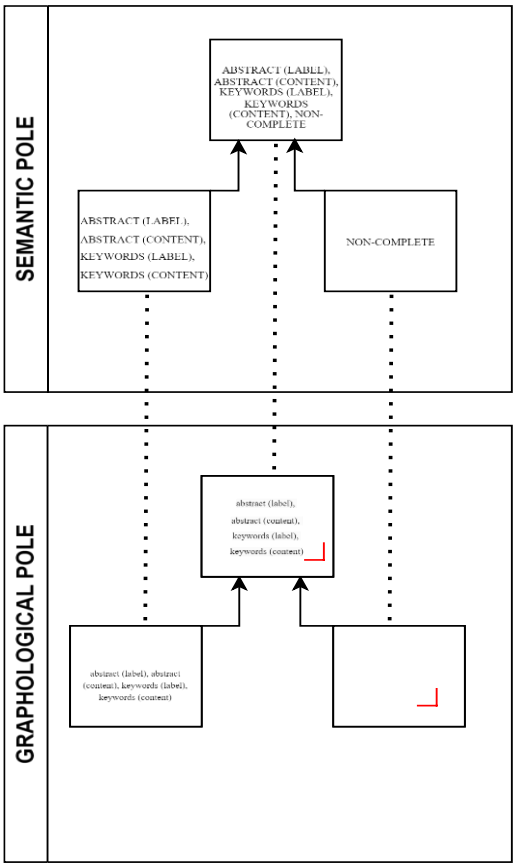


Figure 69. Schematic EEAxC in international journals in law

In linguistics, it was found that the local journals can be characterized by 19 linguistic and 5 paralinguistic constructions. Simultaneously, 6 linguistic and 3 paralinguistic constructions were determined in the international journals (see Table 22). In this case, the regularities of the international journals are equivalent to the ones described in relation to the discipline; therefore, this section concentrates on the regularities identified with reference to the local journals.

	CONSTRUCTIONS (LOCAL)	FREQUENCY PER ALL SEQUENCES (LOCAL)	CONSTRUCTIONS (INTERNATIONAL)	FREQUENCY PER ALL SEQUENCES (INTERNATIONAL)
LINGUISTIC	abstract (content) (Polish)	1/8	abstract (label)	5/6
	abstract (label)	7/8	article (DOI) (content)	1/6
	abstract (label: summary)	1/8	keywords (label)	6/6
	abstract (label: summary) (Polish)	1/8	keywords (content)	6/6
	article (DOI) (content)	1/8	time log (accepted) (label+content)	1/6
	article (title)	4/8	time log (received) (label+content)	1/6
	author (affiliation)	1/8		
	author (contact) (content)	1/8		
	author (name)	1/8		
	issue (page)	1/8		
	issue (page range)	1/8		
	journal (affiliation)	1/8		
	journal (ISSN)	1/8		
	journal (title)	1/8		
	journal (website)	1/8		
	keywords (content)	7/8		
	keywords (label)	7/8		
	keywords (content) (Polish)	1/8		
	keywords (label) (Polish)	1/8		
PARALINGUISTIC	independent	3/8	independent	6/6
	non-independent	5/8	continuous	6/6
	continuous	5/8	non-complete	6/6
	non-continuous	3/8		
	non-complete	8/8		

Table 22. Linguistic and paralinguistic constructions in local and international journals in linguistics

As indicated in Table 22, only one core attribute was identified in local journals, namely non-complete. Therefore, a more schematic construction encompassing the co-occurrence of linguistic and paralinguistic constructions was not established.

In literary studies, in turn, local journals were identified with 13 linguistic and 5 paralinguistic constructions. Simultaneously, international journals were determined to have 8 linguistic and 6 paralinguistic constructions (see Table 23).

	CONSTRUCTIONS (LOCAL)	FREQUENCY PER ALL SEQUENCES (LOCAL)	CONSTRUCTIONS (INTERNATIONAL)	FREQUENCY PER ALL SEQUENCES (INTERNATIONAL)
LINGUISTIC	abstract (label)	6/8	abstract (label)	5/6
	abstract (label: summary)	2/8	article (title)	2/6
	article (DOI) (content)	1/8	author (affiliation)	2/6
	article (DOI) (label)	1/8	author (contact) (content)	2/6
	article (title)	3/8	author (name)	2/6
	author (affiliation)	2/8	author (ORDIC) (content)	1/6
	author (name)	2/8	keywords (content)	6/6
	author (ORDIC) (content)	1/8	keywords (label)	6/6
	author (ORCID) (label)	1/8		
	keywords (content)	6/8		
	keywords (label)	6/8		

	keywords (content) (Polish)	1/8		
	keywords (label) (Polish)	1/8		
PARALINGUISTIC	independent	6/8	independent	4/6
	non-independent	2/8	non-independent	2/6
	continuous	4/4	continuous	4/6
	non-continuous	4/4	non-continuous	2/6
	non-complete	8/8	complete	1/6
			non-complete	5/6

Table 23. Linguistic and paralinguistic constructions in local and international journals in literary studies

As in local journals in linguistics, only one core attribute was identified in local journals in literary studies, namely non-complete. Therefore, a more schematic representation of EEAxC was not established in this sub-community. In the case of international journals, the following core attributes were identified: keywords (label) and keywords (content). However, since no core attributes were established based on the data pertaining to paralinguistic constructions, a more schematic representation of EEAxC was not determined as well.

6.2.2.3. The degree of complexity

When evaluating the degree of complexity, the study indicated that the regularities either align with or diverge from the established ones in the first stage. Moreover, the study indicated that a narrower perspective allows for the identification of new structural invariants. Specifically, additional stable relationships were established in law and literary studies beyond those identified in the first stage. In linguistics, however, identifying structural invariants was only possible at this stage.

To begin with, it was determined that the local journals in law can be described with sequences consisting of 1 to 5 punctuation units (see Table 24).

SEQ.	**P. UNIT 1**	**P. UNIT 2**	**P. UNIT 3**	**P. UNIT 4**	**P. UNIT 5**
1	abstract (label) – abstract (content) | keywords (label) – keywords (content)				
2	article (title)	abstract (label: summary)	abstract (content)	keywords (label) – keywords (content)	
3	author (name) | author (ORCID) (label) | author (affiliation) | author (contact) (content)	article (title)	abstract (label)	abstract (content)	keywords (label) – keywords (content)
4	abstract (label: summary)	abstract (content)	keywords (label) – keywords (content)		
5	abstract (label) – abstract (content	keywords (label) – keywords (content)			

Table 24. Punctuation units in local journals in law

Simultaneously, it was determined that, similar to the discipline in general, the keywords (label) and keywords (content) constructions always co-occur in the same punctuation unit in this sub-community. However, it was also found that the abstract (label: summary) construction always appears alone in a separate punctuation unit.

When it comes to the establishment of the most frequently occurring sequence, it was found that this construction complies with the one presented in Section 6.2.1.3, which was presented in relation to the most common construction in the discipline.

The international journals in law, in turn, can be characterized with reference to sequences consisting of 3 to 6 punctuation units (see Table 25).

SEQ.	P. UNIT 1	P. UNIT 2	P. UNIT 3	P. UNIT 4	P. UNIT 5	P. UNIT 6
6	abstract (label)	abstract (content)	keywords (label) – keywords (content)			
7	journal (title) | issue (number)	article (title)	author (name)	abstract (label)	abstract (content)	keywords (label) – keywords (content)
8	abstract (label)	abstract (content)	keywords (label) – keywords (content)			
9	journal (title) | issue (number) – issue (page range) | article (DOI) (label) – article (DOI) (content)	article (title)	author (name)	abstract (label)	abstract (content)	keywords (label) – keywords (content)

Table 25. Punctuation units in international journals in law

Similarly, the data indicated that the keywords (label) and the keywords (content) constructions consistently appear together in the same punctuation unit. However, it was also determined that the abstract (label) construction and the article (title) construction always appear alone a in separate punctuation unit. Additionally, it was shown that the journal (title) construction co-occurs with the issue (page) construction.

When it comes to the most frequently occurring sequence, it was determined that it is the EEAxC consisting of 3 punctuation units, summarizing the article and providing information about its key terms, and its shape indicates its dependence in relation to other linguistic constructions on the PDF page, non- equal status of all its parts, and the lack of reaching its virtual limit (see Figure 70).

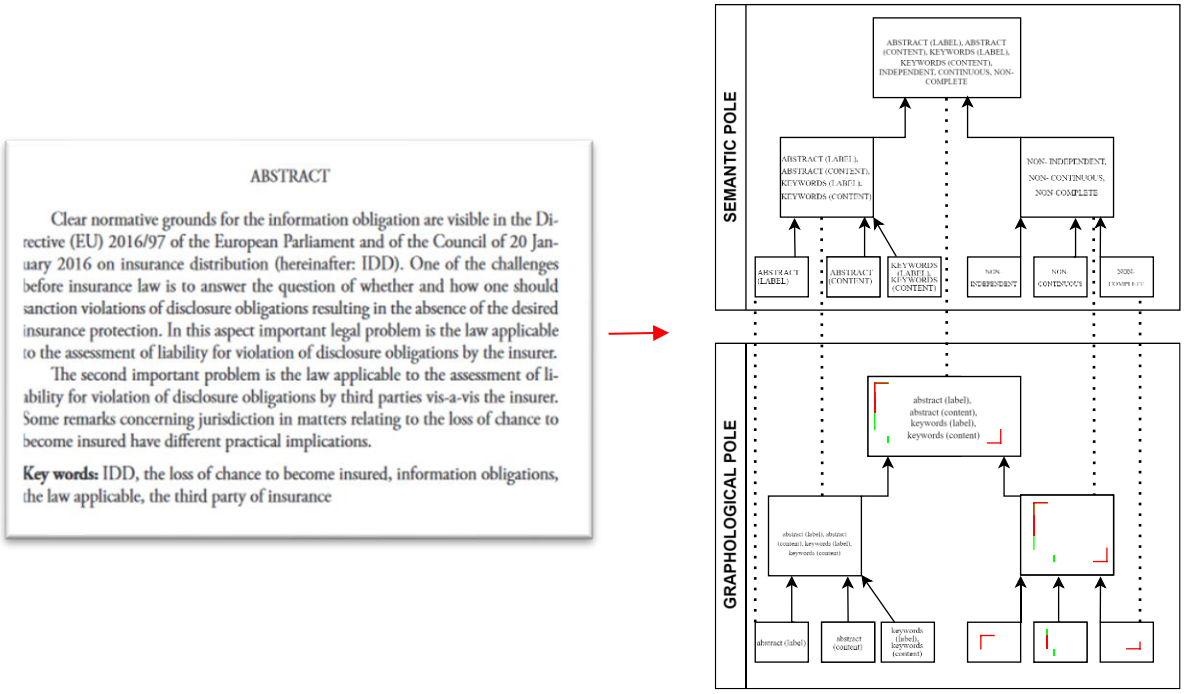


Figure 70. EEAxC in international journals in law

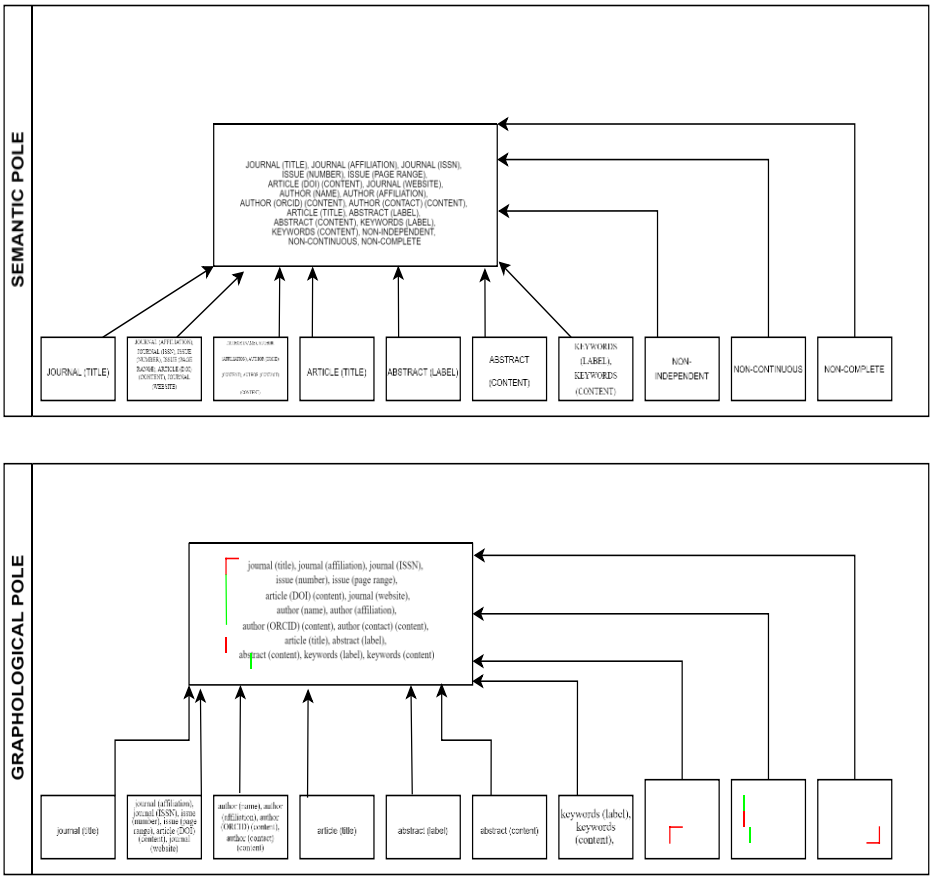
For the local journals in linguistics, it was determined that they can be described in terms of sequences consisting of 2 to 8 punctuation units (see Table 26).

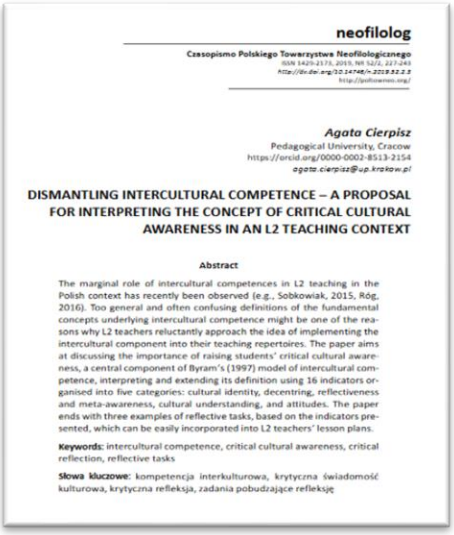
SEQ.	P. UNIT 1	P. UNIT 2	P. UNIT 3	P. UNIT 4	P. UNIT 5	P. UNIT 6	P. UNIT 7	P. UNIT 8
1	journal (title)	journal (affiliation) | journal (ISSN) – issue (number) – issue (page range) | article (DOI) (content) | journal (website)	author (name) | author (affiliation) | author (ORCID) (content) | author (contact) (content)	article (title)	abstract (label)	abstract (content)	keywords (label) – keywords (content	keywords (label) (Polish) – keywords (content) (Polish
2	abstract (label)	article (title)	abstract (content)	keywords (label) – keywords (content)				
3	keywords (label) – keywords (content)	abstract (label) – abstract (content)						
4	article (title)	abstract (label: summary)	abstract (content)	keywords (label) – keywords (content)				
5	abstract (label)	abstract (content)						
6	abstract (label)	abstract (content)						
7	article (title)	issue (page)	abstract (label)	abstract (content)	abstract (label: summary) (Polish)	abstract (content) (Polish)		

8	abstract (label) – abstract (content)	keywords (label) – keywords (content)						

Table 26. Punctuation units in local journals in linguistics

Based on the above table, it was noted that, in this sub-community, the keywords (label) and keywords (content) constructions consistently co-occur in the same punctuation unit, and the article (title) construction always appears alone in a separate punctuation unit. Additionally, it was found that the most frequent sequence consists of 7 punctuation units, i.e. the EEAxC summarizing the article and providing additional information about its key terms in two languages, its title, author(s), identifier, issue and journal in which it is published, and its shape indicates its dependence in relation to other linguistic constructions on the PDF page, non-equal status of all its parts, and the lack of reaching its virtual limit. However, due to the number of punctuation units, the representation of this sequence is provided in a different way, i.e. firstly the representation is given and then a construct is provided below (see Figure 71).





neofilolog

Czasopismo Polskiego Towarzystwa Neofilologicznego

Agata Cierpisz
Pedagogical University, Cracow
https://orcid.org/0000-0002-8513-2154
agata.cierpisz@up.krakow.pl

DISMANTLING INTERCULTURAL COMPETENCE – A PROPOSAL FOR INTERPRETING THE CONCEPT OF CRITICAL CULTURAL AWARENESS IN AN L2 TEACHING CONTEXT

Abstract

The marginal role of intercultural competences in L2 teaching in the Polish context has recently been observed (e.g., Sobkowiak, 2015, Róg, 2016). Too general and often confusing definitions of the fundamental concepts underlying intercultural competence might be one of the reasons why L2 teachers reluctantly approach the idea of implementing the intercultural component into their teaching repertoires. The paper aims at discussing the importance of raising students' critical cultural awareness, a central component of Byram's (1997) model of intercultural competence, interpreting and extending its definition using 16 indicators organised into five categories: cultural identity, decentring, reflectiveness and meta-awareness, cultural understanding, and attitudes. The paper ends with three examples of reflective tasks, based on the indicators presented, which can be easily incorporated into L2 teachers' lesson plans.

Keywords: intercultural competence, critical cultural awareness, critical reflection, reflective tasks

Słowa kluczowe: kompetencja interkulturowa, krytyczna świadomość kulturowa, krytyczna refleksja, zadania pobudzające refleksję

Figure 71. EEAxC in local journals in linguistics

For the international journals in linguistics, the same regularities as the ones determined in the first stage of analysis were delineated. More precisely, the same degree of complexity was identified and no structural invariants were determined. Simultaneously, the same sequence consisting of a single punctuation unit, which was presented in Section 6.2.1.3, was found to be the most frequent in this sub-community.

With reference to the local journals in literary studies, it was established that they can be described in terms of sequences consisting of 1 to 5 punctuation units (see Table 27).

SEQ.	P. UNIT 1	P. UNIT 2	P. UNIT 3	P. UNIT 4	P. UNIT 5
1	abstract (label)	abstract (content)	keywords (label) – keywords (content)		
2	abstract (label) | abstract (content)	keywords (label) – keywords (content)			
3	article (title)	author (name) | author (affiliation)	abstract (label) | abstract (content)	keywords (label) – keywords (content)	
4	abstract (label: summary)	abstract (content)			
5	article (title)	article (DOI) (label) – article (DOI) (content)	author (name) – author (affiliation) | author (ORCID) (label) – author (ORCID) (content)	abstract (label) – abstract (content)	keywords (label) – keywords (content)
6	abstract (label) – abstract (content	keywords (label) – keywords (content)			
7	abstract (label) | abstract (content)				
8	article (title)	abstract (label: summary)	abstract (content)	keywords (label) – keywords (content)	keywords (label) (Polish) – keywords (content) (Polish)

Table 27. Punctuation units in local journals in literary studies

Based on the data above, it was established that, in this subcommunity, the keywords (label) and keywords (content) constructions always appear in separate punctuation units. Moreover, it was shown that the abstract (summary: label) and article (title) always appear alone in separate punctuation units. Additionally, the study found that the author (name) and author (affiliation) constructions always co-occur in the same punctuation unit. At the same time, it was determined that the most frequently occurring sequence is the one consisting of 3 punctuation units, summarizing the article and providing information about its key terms, and its shape indicates its independence in relation to other linguistic constructions on the PDF page, equal status of all its parts, and the lack of reaching its virtual limit (see Figure 72).

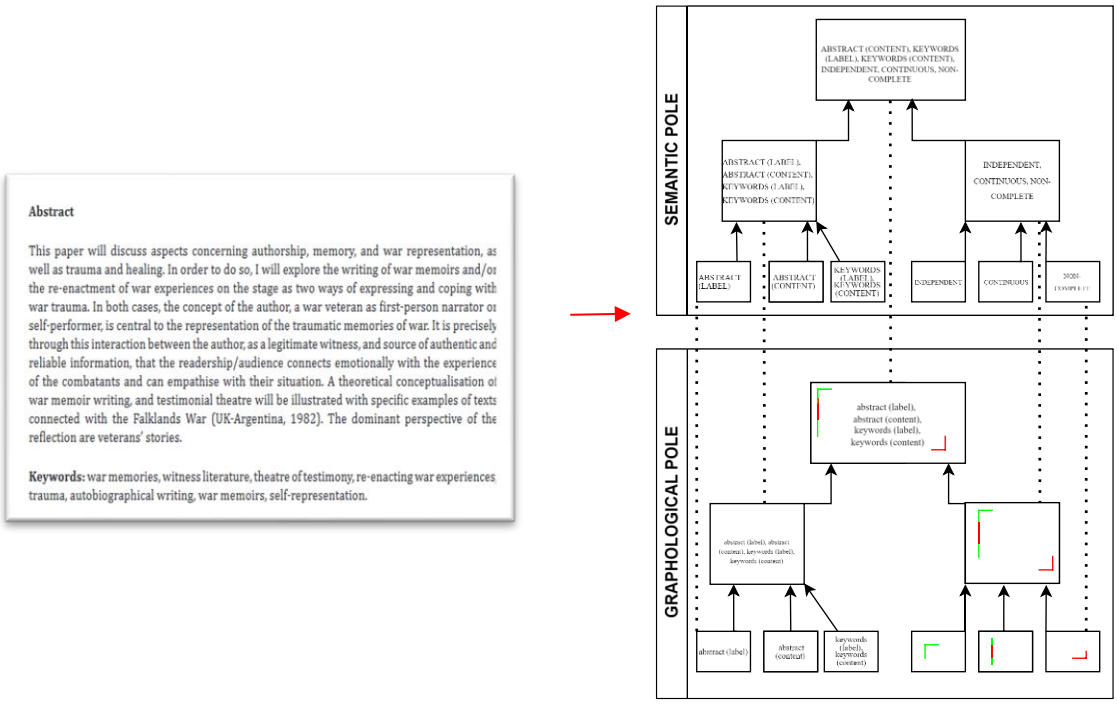


Figure 72. EEAxC in local journals in literary studies

In relation to international journals in literary studies, it was established that they can be characterized based on sequences consisting of 2 to 5 punctuation units (see Table 28).

SEQ.	P. UNIT 1	P. UNIT 2	P. UNIT 3	P. UNIT 4	P. UNIT 5
9	abstract (label) – abstract (content)	keywords (label) – keywords (content)			
10	abstract (label) – abstract (content)	keywords (label) – keywords (content)			
11	abstract (label)	abstract (content)	keywords (label) – keywords (content)		

12	abstract (content)	keywords (label) – keywords (content)			
13	author (name)	author (affiliation) – author (contact) (content) | author (ORCID) (content)	article (title)	abstract (label) – abstract (content)	keywords (label) – keywords (content)
14	author (name)	author (affiliation) – author (contact) (content)	article (title)	abstract (label) – abstract (content)	keywords (label) – keywords (content)

Table 28. Punctuation units in international journals in literary studies

Furthermore, the analysis revealed a consistent pattern within this community concerning the co-occurrence of the keywords (label) and keywords (content) constructions, appearing together in the same punctuation units. Similarly, it was found that the author (affiliation) and author (contact) (content) constructions consistently appear within the same punctuation unit. In contrast, the author (name) and article (title) constructions were observed to appear separately in their own punctuation units. Regarding the most frequently occurring sequence, it was found that it aligns with the one identified during the first stage of the study (see Section 6.2.1.3).

6.3. Findings

This section encapsulates the findings of the present study, addressing the research objectives specified in Section 6.1.1.. First, general conclusions are outlined, which are then followed by a detailed discussion of identified regularities pertaining to the academic disciplinary communities and their sub-communities.

Overall, the analysis indicated that it is possible to identify recurring sequences of linguistic and paralinguistic constructions signaling the emerging nature of the elaborated abstract construction (Research Objective 1). Therefore, it can be assumed that a new discourse genre construction attributable to academic ELF disciplinary communities seems to be developing. At the same time, such a form-meaning pairing arises with its own regularities concerning its degree of fixedness, schematicity and, complexity (Research Objective 2), which are primarily discipline-specific (Research Objective 3) and journal-specific (Research Objective 4).

To be more precise, when it comes to the degree of fixedness, the first stage of the study revealed a universal tendency across disciplines to favor emergent (occurring only once) over emerging (occurring at least twice) sequences indicating the elaborated abstract construction. Subsequently, it was shown that 4 sequences in law, 5 sequences in linguistics, and 4 in literary studies constituted at least 5% of the data. At the same time,

the analysis demonstrated that the degree of fixedness of the emerging constructions varies across disciplines. For instance, the most frequent sequence constituted 9% of the data in law, 15% in linguistics, and 18% in literary studies. Furthermore, it was observed that, among the emerging sequences, the majority are discipline-specific, with only one sequence deemed universal as no statistical differences were identified. However, universal sequences were also identified for linguistics and literary studies, as well as for law and literary studies, concurrently showing that all universal sequences can be characterized by means of the same paralinguistic constructions, i.e. "independent, continuous, non-complete".

When it comes to the degree of schematicity, it was shown that these disciplines differ in terms of more schematic constructions, which can be determined based on the identified sequences. In this way, core attributes for each discipline were also determined, which included 3 linguistic and 1 paralinguistic constructions in law, 2 linguistic and 3 paralinguistic constructions in linguistics, and 3 linguistic and 2 paralinguistic constructions in literary studies.

Lastly, the analysis showed that disciplines also differ in terms of the degree of complexity. In law, for example, sequences consisting of 1 to 5 punctuation units were identified. In linguistics, sequences consisting of 1 to 4 units were determined. In literary studies, in turn, sequences consisting of 2 to 3 punctuation units were identified. At the same time, it was indicated that the most frequent sequence in law and linguistics comprises one punctuation unit whereas in literary studies it comprises two punctuation units. Moreover, it was shown that it is possible to determine structural invariants among attributes in law and literary studies. For linguistics, however, no structural invariants were identified. The results are summarized in Table 29 below.

	LAW	**LINGUISTICS**	**LITERARY STUDIES**
PREFERENCE FOR EMERGENT/EMERGING SEQUENCES	emergent	emergent	emergent
NUMBER OF ANALYZED SEQUENCES	4	5	3
CORE ATTRIBUTES	keywords (label), keywords (content), abstract (label_X), non-complete	keywords (label), keywords (content), independent, continuous, non-complete	abstract (label), keywords (label), keywords (content), independent, continuous,
THE NUMBER OF PUNCTUATION UNITS	1-5	1-4	2-3
STRUCTURAL INVARIANTS	yes	no	yes
THE MOST FREQUENT SEQUENCE	abstract (label) – abstract (content) | keywords (label) – keywords (content), independent, continuous, non-complete	abstract (label) | abstract (content) | keywords (label) – keywords (content), independent, continuous, non-complete	abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, complete

Table 29. Disciplines: findings

At the same time, conclusions drawn from the second part of the study indicated that regularities established in the first stage of the analysis mostly reflect tendencies of a larger sub-community within the discipline. In other words, a more fine-grained perspective revealed new tendencies specific to the smaller sub-community.

Firstly, concerning the degree of fixedness, it was demonstrated that disciplinary preferences regarding emergent sequences are the same in two types of sub-communities only in the case of law. In linguistics, for instance, emerging sequences were found to be more common in international journals. At the same time, no statistical differences were identified between journals in literary studies. Among the emerging sequences, 5 (local) and 4 (international) sequences were identified in law, 8 (local) and 6 (international) in linguistics, and 8 (local) and 6 (international) in literary studies. Furthermore, quantitative data indicated that the most frequent sequences in international journals in each discipline exhibit a higher degree of fixedness relative to disciplinary norms. For example, at the disciplinary level, the discipline of literary studies was characterized by the most frequent sequence constituting 18% of the data, whereas in international journals, the most frequent sequence constituted 54%. Regarding universality, however, tendencies emerging from the analysis of more local and international journals in law and linguistics mirror the tendencies identified at the disciplinary level. In the case of literary studies, however, contrary to what was established in the first stage, the majority of sequences were identified as universally applied in the discipline.

When it comes to the degree of schematicity, it was possible to determine more schematic construction in two types of journals only in law. In the case of local journals, it coincided with the one established at the disciplinary level, while in international journals, it was only slightly modified. In linguistics, a more schematic construction was not specified in local journals, but the one emerging in international journals coincided with the one specified at the disciplinary level in global journals. Similarly, in literary studies, a more schematic construction was not specified for local journals; however, it was also not specified for global journals. As a consequence, different core attributes were determined (i.e. 3 linguistic and 1 paralinguistic in local and global journals in law, one paralinguistic in local journals in linguistics and 2 linguistic and 3 paralinguistic in global journals, and 1 paralinguistic in local journals in literary studies and two linguistic in global journals).

Finally, regarding degrees of complexity, constructions in global journals in law were characterized as more complex than in local journals (ranging from 3 to 6 and from 1 to 5 punctuation units, respectively). In linguistics, on the other hand, constructions in local journals were found to be more complex than in global journals (ranging from 2 to 8 and from 1 to 4 punctuation units, respectively). In literary studies, constructions are roughly similar (i.e., ranging from 1 to 5 and from 2 to 5 punctuation units in local and global journals, respectively). Simultaneously, it was established that the most common sequence in both local and international law journals consists of a sequence comprising 3 punctuation units. In linguistics, a construction comprising 8 punctuation units was identified in local journals and 1 in global journals. In literary studies, a construction consisting of 3 punctuation units was identified in local journals and of 2 in global journals. Structural invariants, apart from constructions in global journals in linguistics, were identified in each sub-community. The results are summarized in Table 30.

	LAW		**LINGUISTICS**		**LITERARY STUDIES**	
	L	**I**	**L**	**I**	**L**	**I**
PREFERENCE FOR EMERGENT/EMERGING SEQUENCES	emergent	emergent	emergent	emerging	emergent /emerging	emergent /emerging
NUMBER OF ANALYZED SEQUENCES	5	4	8	6	8	6
CORE ATTRIBUTES	keywords (label), keywords (content), abstract (label_X), non-complete	keywords (label), keywords (content), abstract (label), and non-complete.	non-complete	keywords (label), keywords (content), independent, continuous, non-complete	non-complete	keywords (label), keywords (content),
THE NUMBER OF PUNCTUATION UNITS	1–5	3–6	2–8	1–4	1–5	2–5
STRUCTURAL INVARIANTS	yes	yes	yes	no	yes	yes
THE MOST FREQUENT SEQUENCE	abstract (label: summary) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete	abstract (label) || abstract (content) || keywords (label) – keywords (content), non-independent, non-continuous, non-complete	journal (title) || journal (affiliation) | journal (ISSN) – issue (number) – issue (page range) | article (DOI) (content) | journal (website) || author (name) | author (affiliation) | author (ORCID) (content) | author (contact) (content) || article (title) || abstract	abstract (label) | abstract (content) | keywords (label) – keywords (content), independent, continuous, non-complete	abstract (label) || abstract (content) || keywords (label) – keywords (content), independent, continuous, non-complete	abstract (label) – abstract (content) || keywords (label) – keywords (content), independent, continuous, complete

			(label) || abstract (content) || keywords (label) – keywords (content) || keywords (label) (Polish) – keywords (content) (Polish), non-independent, non-continuous, non-complete			

Table 30. Journals: findings

6.4. Limitations

However, it is worth noting that the findings of the study were shaped by several constraints, which can be roughly categorized as theoretical, methodological, and practical. When it comes to theoretical constraints, considering the emphasized need for a precise delineation of the communities under study and a careful identification of their tendencies, it should be noted that the findings of this study may be perceived as limited to delineating tendencies within academic disciplinary ELF communities and their sub-communities only with reference to Polish journals in three disciplines publishing articles in English from 2018 to 2021.

As for methodological constraints, it should be noted that this study utilizes the data sourced from the DISCOWER corpus, which was developed using an innovative method for determining spatial groupings through Gestalt-based schematic structures, which clearly requires further research (and potential refinements). Therefore, it is important to underscore that the findings discussed in this chapter should be viewed as tentative tendencies that necessitate further investigation.

At the same time, the proposed methodology, although it is based on related studies, is perceived as just one of the possible ways to operationalize discourse genre constructions. In other words, considering the initial stage of research into co-occurring linguistic and paralinguistic constructions, it seems likely that new solutions will emerge that could deepen the insights derived from this analysis.

Finally, concerning practical limitations, the assumption that the study focuses on only sequences comprising at least 5% of the data undoubtedly resulted in establishing tendencies pertaining to only a part of a fuller picture that could be obtained by utilizing

the data from the DISCOWER corpus. Nevertheless, at this stage, adopting such an approach was perceived as an attempt to develop a solution that would make this study manageable.

Conclusion

As indicated at the beginning of this dissertation, its main aim was to identify and explore the emerging elaborated abstract construction within three academic disciplinary communities, as such research was perceived as potentially contributing to further advancements in two fields, i.e. academic English as a lingua franca and Cognitive (Construction) Grammar. While the validity of the above-delineated aim was gradually reinforced by specific gaps identified in the first two parts of the dissertation, it was directly addressed in the third part, where the study, constituting the heart of this project, was presented.

When it comes to academic English as a lingua franca, the method of collecting and analyzing data proposed in this study allows for integrating two currently (under)developed approaches, i.e. Linear Unit Grammar and Construction Grammar, in academic ELF research. More specifically, it allows for establishing ELF-specific constructions based in Gestalt-based cognitive processes, thereby facilitating the extension of research within LUG concentrated on spoken data, and contributing to the ongoing development of construction-based approaches, which are still said to be in their infancy.

Simultaneously, the findings of the study appear to complement the delineated theoretical assumptions and empirical evidence regarding academic ELF. In other words, a comprehensive understanding of academic ELF is achieved by considering both emergent and emerging units and their universal and relative nature. However, due to the primary focus on emerging units adopted in this study, its findings particularly deepen the perspective that it is possible to identify stabilizing units within academic disciplinary communities, which are predominantly community-specific.

When it comes to Cognitive (Construction) Grammar, in turn, the proposed method of collecting and analyzing data seems to advance current research on discourse genre constructions by highlighting the role of paralinguistic units and their interplay with linguistic units. Furthermore, this study offers a holistic description of academic discourse genre constructions, taking into account recurring sequences of lower-level constructions, which seems to be currently overlooked in the related research. Concurrently, the

conclusions drawn from this study particularly confirm the need to explore discourse genre constructions while considering potential differences between communities.

On the other hand, identifying the emerging elaborated abstract construction and determining disciplinary tendencies can be perceived as beneficial for other research fields, such as English for Academic Purposes, as the findings of this study may contribute to the understanding of evolving practices in (ELF) research dissemination. Moreover, the creation of a freely available written ELF corpus undoubtedly constitutes a valuable database that can be utilized not only for further construction-based ELF-related studies but also for further explorations from various research perspectives.

References

Abrar-ul-Hassan, Shahid. 2021. “Linguistic Capital in the University and the Hegemony of English: Medieval Origins and Future Directions,” *Sage Open* 11(2), 147–154.

Adamou, Evangelia, and Yaron Matras. 2021. *The Routledge Handbook of Language Contact*. New York: Routledge.

Ädel, Annelie. 2020. “Corpus Compilation,” In Magali Paquot and Stefan Th. Gries (eds.), *A Practical Handbook of Corpus Linguistics*, 3–24.

Al-Jasser, Jasser. 2012. “Pidginization Theory and Second Language Learning/Acquisition,” *Journal of King Saud University – Languages and Translation* 24.2, 71–74.

Alptekin, Cem. 2011. “Beyond ENL Norms in ELF Use: A Cognitive Perspective on EFL Output,” In Carsten Roever and Neomy Storch (eds.), *Australian Review of Applied Linguistics* 34(2), 148–164.

Alptekin, Cem. 2013. “English as a Lingua Franca through a Usage-Based Perspective: Merging the Social and the Cognitive in Language Use,” *Language, Culture and Curriculum* 26.2, 197–207.

Ammon, Ulrich. 2000. “Towards More Fairness in International English: Linguistic Rights of Non-native Speakers?,” In Robert Phillipson (ed.), *Rights to Language Equity, Power, and Education* 1, 111–116.

Anderson, Benedict. 1991. *Imagined. Communities: Reflections on the Origin and Spread of Nationalism.* New York: Verso.

Ang, Hoon. 1997. “Chinese Consumers’ Perception of Alpha-numeric Brand Names,” *Journal of Consumer Marketing* 14, 220–233.

Antonopolou, Eleli, and Kiki Nikiforidou. 2011. “Construction Grammar and Conventional Discourse: A Construction-Based Approach to Discoursal Incongruity,” Journal of Pragmatics 43, 2594–2609.

Auer, Peter and Stefan Pfäender. 2011. “Constructions: Emergent or Emerging?,” In Peter Auer, Gesa von Essen, and Werner Frick (eds.), *Constructions: Emerging and Emergent*, 1–21.

Back, Andrea. 2015. *English as a Lingua Franca: The Case for ELF as an Independent, Natural and Legitimate Lingua Franca*. Venice: Ca’ Foscari University of Venice.

Baicchi, Annalisa, and Aneider Iza Erviti. 2018. “Genre as Cognitive Construction,” *Pragmatics and Cognition* 25 (3), 576–601.

Baird, Robert, Will Baker, and Mariko Kitazawa. 2014. “The Complexity of English as a Lingua Franca,” In Barbara Seidlhofer and Anna Mauranen (eds.), *Journal of English As A Lingua Franca* 3, 171–196.

Bakker, Peter. 1995. “Pidgins,” In Jacques Arends, Pieter Muysken, and Norval Smith (eds.), *Pidgins and Creoles: An Introduction*, 25–39.

Bakker, Peter. 2009. “Pidgins versus Creoles and Pidgincreoles,” In Silvia Kouwenberg and John V. Singler (eds.), *The Handbook of Pidgin and Creole Studies*, 130–157.

Bakker, Peter, and Yaron Matras (eds.). 2013. *Contact Languages. A Comprehensive Guide*. Boston/Berlin: Walter de Gruyter.

Barlow, Michael and Suzanne Kemmer. 2000. *Usage Based Models of Language*. Stanford: Center for the Study of Language and Information.

Barsalou, Lawrence W. 1999. “Perceptual Symbol Systems,” *Behavioral and Brain Sciences* 22, 577–660.

Barsalou, Lawrence W. 2009. “Simulation, Situated Conceptualization, and Prediction,” *Philosophical Transactions of The Royal Society B: Biological Sciences* 364(1521), 1281–1289.

Barsalou, Lawrence W. 2012. “The Human Conceptual System,” In Michael Spivey, Ken McRae, Marc Joanisse (eds.), *The Cambridge Handbook of Psycholinguistics*.

Barsalou, Lawrence W. 2017. “What Does Semantic Tiling of the Cortex Tell Us about Semantics?,” *Neuropsychologia* 105, 18–38.

Barsalou, Lawrence W. 2020. “Challenges and Opportunities for Grounding Cognition,” *Journal of Cognition* 3 (1), 1–24.

Barsalou, Lawrence W. and Dorrit Billman. 1989. “Systematicity and Semantic Ambiguity,” In David S. Gorfein (ed.), *Resolving Semantic Ambiguity*, 146–203.

Barsalou, Lawrence W., Barbara J. Luka, Karen L. Olseth and Wenchi Yeh. 1993. “Concepts and Meaning,” *Chicago Linguistic Society* 29, 23–61.

Bateman, John, Judy Delin, and Renate Henschel. 2007. “Mapping the Multimodal Genres of Traditional and Electronic Newspapers,” *New Directions in the Analysis of Multimodal Discourse*, 147–172.

Bayyurt, Yasemin, Yavuz Kurt, Elifcan Öztekin, Luis Guerra, Lili Cavalheiro, and Ricardo 2019. “English Language Teachers Awareness of English as a Lingua

Franca in Multilingual and Multicultural Contexts," *Eurasian Journal of Applied Linguistics* 5.2, 185–202.

Becher, Tony, and Paul R. Trowler. 2006. *Academic Tribes and Territories – Intellectual Enquiry and the Culture of Disciplines*. Buckingham: The Society for Research into Higher Education & Open University Press.

Behrens, Heike. 2009. "Usage-based and Emergentist Approaches to Language Acquisition," *Linguistics* 47, 383–411.

Bennett, Phillip, Maria Tassinari, Ena Hollinshead, and Fergal Bradley. 2020. "Reflections on the Editorial Process of an Academic Journal," *Relay Journal* 3(1), 100–109.

Bergs, Alexander. 2008. "Can we Take Construction Grammar Beyond Sneezing Napkins off Tables?," Anglistentag and Klaus Stierstorfer (eds.), *Proceedings / Anglistentag 2007 Münster*, 269–276.

Bernardini, Marco and Silvia Baroni. 2004. "BootCaT: Bootstrapping Corpora and Terms from the Web," Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC'04), Lisbon, Portugal. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2004/pdf/509.pdf

Bhatia, Vijay K. 1993. *Analysing Genre: Language Use in Professional Settings*. London /New York: Routledge.

Björkman, Beyza. 2008. "English as the Lingua Franca of Engineering Education," *Nordic Journal of English Studies* 7.3, 103–122.

Björkman, Beyza. 2013. *English as an Academic Lingua Franca*. Boston / Berlin: de Gruyter.

Blumenthal-Dramé, Alice. 2012. *Entrenchment in Usage-Based Theories: What Corpus Data Do and Do Not Reveal About The Mind*. Boston / Berlin: De Gruyter Mouton.

Boers, Frank. 1996. *Spatial Prepositions and Metaphor: A Cognitive-semantic Journey along the UP-DOWN and the FRONT-BACK Dimensions*. Tübingen: Gunter Narr.

Bohát, Róbert. 2018. "Corpus & Discovery Learning in High-School Academic English and Academic Writing: COHAT and Other Corpora," *The Foundations and Versatility of English Language Teaching (ELT)*, 57–74.

Bonato, Sarah. 2018. *Searching the Grey Literature. A Handbook for Researching Reports, Working Papers, and Other Unpublished Research*. Lanham / Boulder: Rowman & Littlefield.

Borghi, Anna, and Ferdinand Binkofski. 2014. *Words as Social Tools: An Embodied View on Abstract Concepts*. New York/Heidelberg: Springer.

Breiteneder, Angelika, Marie-Luise Pitzl, Stefan Majewski, and Theresa Klimpfinger. 2006. "VOICE Recording – Methodological Challenges in the Compilation of a Corpus of Spoken ELF," *Nordic Journal of English Studies* 5(2), 161–187.

Breyer, Themo, Stefan Pfänder and Oliver Ehmer. 2011. "Improvisation, temporality and emergent constructions," In Peter Auer nd Stefan Pfänder (eds.), *Constructions Emerging and Emergent*, 186–217.

Brideau, Kate, and Hand Charles Berret. 2014. "A Brief Introduction to Impact: 'The Meme Font'," *Journal of Visual Culture* 13(3), 307–313. Thousand Oaks: SAGE Publishing.

Brosch, Cyril. 2015. "On the Conceptual History of the Term Lingua Franca," *Apples – Journal of Applied Language Studies* 9.1, 71–85.

Brunner, Marie-Louise. 2022. *Understanding Intercultural Communication: Negotiating Meaning and Identities in English as a Lingua Franca Skype Conversations*. Berlin: Springer Berlin Heidelberg.

Butler, Christopher, and Francisco Gonzálvez-García. 2014. *Exploring Functional-Cognitive Space*. Amsterdam: John Benjamins.

Bybee, Joan. 2010. *Language, Usage and Cognition.* Cambridge: Cambridge University Press.

Bybee, Joan. 2023. "What Is Usage-Based Linguistics?," In Manuel Díaz-Campos, Sonia Balasch (eds.), *The Handbook of Usage-Based Linguistics*, 7–29.

Canagarajah, Suresh. 2007. "Lingua Franca English, Multilingual Communities, and Language Acquisition," *The Modern Language Journal* 91, 923–239.

Canagarajah, Suresh. 2018. "Translingual Practice as Spatial Repertoires: Expanding the Paradigm beyond Structuralist Orientations," *Applied Linguistics* 39, 31–54.

Canagarajah, Suresh. 2020. "English as a Resource in a Communicative Assemblage: A Perspective from Flat Ontology," In Christopher J. Hall and Rachel Wicaksono (eds.), *Ontologies of English: Conceptualising the Language for Learning, Teaching, and Assessment*, 295–314.

Capozza, Sarah. 2023. *The Correlation of Business English as Lingua Franca and the Job Satisfaction of Non-native Speakers of English*. Munich: Grin Verlag.

Carey, Ray. 2013. "On the Other Side: Formulaic Organizing Chunks in Spoken and Written Academic ELF," *Journal of English as a Lingua Franca* 2(2), 207–228.

Carey, Ray. 2015. https://www.helsinki.fi/en/researchgroups/english-as-a-lingua-franca-in-academic-settings/research/wrelfa-corpus

Chafe, Wallace L. 1980. *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*. Norwood, NJ: Ablex Publishing Corporation.

Chafe, Wallace L. 1987a. "Cognitive Constraints on Information Flow," In Russell S. Tomlin (ed.), *Coherence and Grounding in Discourse*, 21–51.

Chafe, Wallace L. 1987b. "Prosodic and Functional Units of Language," In Joan Bybee and Larry Hyman (eds.), *Typological Studies in Language* 12, 33–44.

Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. Chicago: University of Chicago Press.

Chen, Xiang. 2001. "Perceptual Symbols and Taxonomy Comparisons." *Philosophy of Science* 68, 8200–8212.

Chen, Xiang. 2003. "Object and Event Concepts: A Cognitive Mechanism of Incommensurability." *Philosophy of Science* 70.5, 962–974.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.

Chomsky, Noam. 1988. *Language and Problems of Knowledge: The Managua Lectures*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2013. "Problems of Projection," *Lingua* 130, 33–49.

Christiansen, Morten, and Nick Chater. 2016. "The Now-or-Never Bottleneck: A Fundamental Constraint on Language," *Behavioral and Brain Sciences* 62, 1–72.

Cienki, Alan. 2015. "Spoken Language Usage Events," *Language and Cognition* 7. 499–514.

Cody, Ron. 2019. *A Gentle Introduction to Statistics Using SAS Studio*. NC: SAS Institute Inc.

Cogo, Alessia. “English as a Lingua Franca: Concepts, Use, and Implications,” *ELT Journal* 66(1), 97–105.

Coleman, James A. 2006. “English-medium Teaching in European Higher Education,” *Language Teaching* 39(1), 1–14.

Coussé, Evie, and Ferdinand von Mengden. 2014. *Usage-based Approaches to Language Change*. Amsterdam / Philadelphia: John Benjamins Publishing.

Cowie, Fiona. 2008. “Innateness and Language,” In Edward N. Zalta (ed.), The Stanford Encyclopedia of Philosophy (Fall 2017 Edition), <https://plato.stanford.edu/archives/fall2017/entries/innateness-language/>.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford: Oxford University Press.

Croft, William and Alan Cruse. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Croft, William. 2009. “Toward a Social Cognitive Linguistics,” In Vyvyan Evans and Stephanie Pourcel (eds.), *New Directions in Cognitive Linguistics*, 1–29.

Crowley. Terry. 2009. “Pidgin and Creole Morphology,” In Silvia Kouwenberg and John V. Singler (eds.), *The Handbook of Pidgin and Creole Studies*, 74–97.

Dahl, Alva. 2018. “The Graphic and Grammatical Structure of Written Texts,” *Studia Neophilologica* 90, 24–36.

Day, Richard R. 2019. “Early Pidginization in Hawaii,” In Glenn Gilbert (ed.), *Pidgin and Creole Languages: Essays in Memory of John E. Reinecke*, 183–200.

Davidse, Kristin, Ngum Meyuhnsi Njende, and Lobke Ghesquière. 2017. “Precious Few and Practically All. A Cognitive Grammar Approach to the Modification of Quantifiers,” *Leuven Working Papers in Linguistics* 6, 34–64.

Davies, Alan. 1991. *The Native Speaker in Applied Linguistics*. Edinburgh: Edinburgh University Press.

Davies, Alan. 2003. *The Native Speaker: Myth and Reality* (2nd edition). Clevedon: Multilingual Matters.

Dąbrowska, Ewa. 2008. “Questions with Long-distance Dependencies: A Usage-based Perspective,” *Cognitive Linguistics* 19(3), 391–425.

Dąbrowska, Ewa, and Dagmar Divjak. 2015. *Handbook of Cognitive Linguistics*. Berlin/Boston: Walter de Gruyter.

Deane, Paul. 1993. *Multimodal Spatial Representation: On the Semantic Unity of Over and Other Polysemous Prepositions*. Essen: Linguistic Agency University of Duisburg.

Dewey, Martin, and Jennifer Jenkins. 2010. "English as a Lingua Franca in the Global Context: Interconnectedness, Variation and Change," *Contending with Globalization in World Englishes*, 72–86. Bristol: Multilingual Matters.

Dewey, Martin, and Pineda Inmaculada. 2020. "ELF and Teacher Education: Attitudes and Beliefs," *ELT Journal* 74(4), 428–444.

Diessel, Holger. 2015. "Usage-based Construction Grammar," in Ewa Dabrowska and Dagmar Divjak (eds.) *Handbook of Cognitive Linguistics* 39, 296–322.

Diessel, Holger. 2017. "Usage-Based Linguistics," In Mark Aronoff (ed.), *Oxford Research Encyclopedia of Linguistics*, http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e- 363?rskey=ivWwgv&result=2.

Dirven, René. 2005. "Major Strands in Cognitive Linguistics," In Francisco J. Ruiz de Mendoza Ibáñez and M. Sandra Peña Cervel (eds.), *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*, 17–68.

Divjak, Dagmar. 2019. *Frequency in Language: Memory, Attention, and Learning*. Cambridge: Cambridge University Press.

Dolberg, Florian. 2019. *Agreement in Language Contact: Gender Development in the Anglo-Saxon Chronicle*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Dove, Guy, Laura Barca, Luca, Tummolini, Anna Borghi. 2020. "Words Have a Weight: Language as a Source of Inner Grounding and Flexibility in Abstract Concepts," *Psychological Research* 1, 1–17.

Dumitru, Magda L. 2016. "Gestalt-like Representations Hijack Chunk-and-Pass Processing," *Behavioral and Brain Sciences* 39, 69.

Edwards, Ray. 2012. *Scientific Gestalt*. Milton Keynes: Author House.

Ehrenstein, Walter H., Lothar Spillman and Vikor Sarris. 2003. "Gestalt Issues in Modern Neuroscience," *Axiomathes* 13: 433–458.

Ellis, Nick, Ute Römer-Barron, and Matthew Brook O'Donnell. 2016. *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. London: Blackwell Publishing.

Enghels, Renata and Mar Garachana Camarero. 2021. “Grammaticalization, Lexicalization, and Constructionalization,” *The Routledge Handbook of Cognitive Linguistics*, 169–188.

Evans, Vyvyan. 2009. *How Words Mean: Lexical Concepts, Cognitive Models, and Meaning Construction*. Oxford: Oxford University Press.

Evans, Vyvyan. 2012. “Cognitive Linguistics,” *WIREs Cognitive Science* 3(2), 129–141.

Evans, Vyvyan, and Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Fauziati, Endang. 2014 https://journal.uad.ac.id/index.php/ADJES/article/view/1680/1177

Fischer, Kerstin and Morgan Aarestrup. 2021. “Relationships Between Construction Grammar(s) and Genre: Evidence from an Analysis of Instagram Posts,” *Journal of Pragmatics* 183, 87–104.

Fiedler, Sabine, and Cyril R. Brosch. 2022. *Esperanto – Lingua Franca and Language Community*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Fiedler, Sabine. 2011. “English as a Lingua Franca – a Native-Culture-Free Code? Language of Communication vs. Language of Identification,” *Apples: Journal of Applied Language Studies* 5, 79–97.

Fillmore, Charles J. 1982. “Frame Semantics,” *Linguistics in the Morning Calm, Selected Papers from SICOL-1981*, 111–137.

Fisher, Kerstin, and Morgan Aarestrup. 2021. “Relationship between Construction Grammar(s) and Genre: Evidence from an Analysis of Instagram Posts,” Journal of Pragmatics, 87–104.

Flowerdew, John. 2022. “Models of English for Research Publication Purposes,” *World Englishes* 41(4), 571–583.

Fonteyn, Lauren. 2019. “A Corpus-based View on the (Aspectual-)Semantics of Modern English Nominalizations,” *Language Sciences*, 77–90.

Forceville, Charles. 2009. “Non-verbal and Multimodal Metaphor in a Cognitivist Framework: Agendas for Research,” In Charles J. Forceville and Eduardo Urios-Aparisi (eds.), *Multimodal Metaphor*, 19–44.

Formentelli, Maicol. 2017. *Taking Stance in English as a Lingua Franca: Managing Interpersonal Relations in Academic Lectures*. Cambridge: Cambridge University Press.

Frederiksen, Diana. 2014. *BELF in the Workplace: A Linguistic Ethnographic Study : An Observation of English as a Lingua Franca Used by Employees at a Swedish Company (Dissertation).*

Frick, Tim, and Kate Eyler-Werve. 2014. *Return on engagement: Content strategy and web design techniques for digital marketing: Second edition*. London / New York: Routledge.

Gallagher, Shaun. 2011. "Interpretations of embodied cognition," In W. Tschacher and C. Bergomi (eds.), *The Implications of Embodiment: Cognition and Communication*, 59–71.

Gallese, Vittorio, and George Lakoff. 2005. "The Brain's Concepts," *Cognitive Neuropsychology* 22, 455–479.

Gardner, David, and Ken Lau. 2018. "Learning and Using English: What Experience and Expectations Do Non-local Asian International Students Bring to their University Studies in Hong Kong?," *Oxford Review of Education*, 1–16.

Gärdenfors, Peter. 2020. "Primary Cognitive Categories Are Determined by Their Invariances," *Frontiers in Psychology* 11.

Gärtner, Hans-Martin. 2014. *Generalized Transformations and Beyond*. Berlin: Akademie Verlag.

Gibbons, Alison. 2012. *Multimodality, Cognition, and Experimental Literature.* London /New York: Routledge.

Gibbons, Alison, and Sara Whiteley. 2018. *Contemporary Stylistics: Language, Cognition, Interpretation.* Edinburgh University Press.

Gibbs, Raymond W. 2017. "Embodiment," In Barbara Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*. Cambridge: Cambridge University Press, 449–462.

Giesteira, Bruno and Diana Tavares. 2010. "Gestalt and Computational Perceptual Approach: Brain Responses Tendencies Given by Visual and Auditory Basic Stimuli," *COGNITIVE 2010: The Second International Conference on Advanced Cognitive Technologies and Applications.*

Gobet, Fernand. 2017. "Entrenchment, Gestalt Formation, and Chunking," in Hans-Jörg Schmid (ed.), *Entrenchment and the Psychology of Language Learning: How we Reorganize and Adapt Linguistic Knowledge*, 245–268.

Goddard, Ives. 1997. "Pidgin Delaware," In Sarah G. Thomason (ed.), *Contact Languages: A Wider Perspective*, 43–98.

Gómez, Verónica. 2020. *English as a Lingua Franca in the Business Domain: The case of University Spin-off Companies in Galicia*. A Coruña: The University of A Coruña.

Godwin-Jones, Robert. 2020. "Towards Transculturality: English as a Lingua Franca in Intercultural Communication and in Online Language Learning," *Languages and International Studies* 23, 1–34.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Goldberg, Adele. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. USA: Princeton University Press.

Gotti, Maurizio. 2021. "Recent Developments Concerning the Use of English for Teaching and Research Purposes," *Language Learning in Higher Education*, 10(2), 287–300.

Graddol, David. 1997. *The Future of English?* London: British Council.

Gramley, Stephan, and Kurt-Michael Pátzold. 2003. *A Survey of Modern English*. London /New York: Routledge.

Grazzi, Enrico. 2020. "Challenging the Interlanguage Hypothesis: The Convergence of EFL and ELF in the English Classroom," *Lingue e Linguaggi* 38, 277–294.

Grazzi, Enrico, and Lucilla Lopriore. 2020. "ELF Awareness for Teacher Education in Italy: Attitudes and actions," *Estudos Linguísticos e Literários* 1(65), 69–89.

Grenfell, Michael, and Vee Harris. 2002. *Modern Languages and Learning Strategies In Theory and Practice.* London / New York: Routledge.

Grenoble, Lenore, and Lindsay Whaley. 2005. *Saving Languages: An Introduction to Language Revitalization*. Cambridge: Cambridge University Press.

Groom, Chloe. 2012. "Non-Native Attitudes Towards Teaching English as a Lingua Franca in Europe," *English Today* 28(1), 50–57.

Guinda, Carmen Sancho. 2022. "Scientific Stylisation or the 'Democracy Dilemma' of Graphical Abstracts," *Publications* 10(1), 1–22.

Günther, Franziska. 2016 "Attention, 'Ception' and Language: Basic Considerations," *Constructions in Cognitive Contexts: Why Individuals Matter in Linguistic Relativity Research* 299, 83–105.

Habjan, Ana Monika. 2016. *Chunking, Emergence, and Online Production as Theoretical Concepts in ELF*. Rome: Roma TrE-Press.

Hall, Joan. 2016. "A Usage-based Account of Multi-competence," In Vivian Cook and Li Wei (eds.), *The Cambridge Handbook of Linguistic Multi-competence*, 183–205.

Hall, Christopher J. 2017. "Cognitive Perspectives on English as a Lingua Franca," In Jennifer Jenkins, Will Baker, Martin Dewey (eds.), T*he Routledge Handbook of English as a Lingua Franca*, 74–84.

Hall, Christopher J. 2018. "The Psycholinguistics of English," *The Routledge Handbook of English Language Studies*, 406–419.

Hart, Christopher. 2016. "The Visual Basis of Linguistic Meaning and Its Implications for Critical Discourse Studies: Integrating Cognitive Linguistic and Multimodal Methods," *Discourse & Society* 27, 335–350.

Hart, Christopher and Marmol Queralto, Javier. 2021. "What Can Cognitive Linguistics Tell Us About Language-Image Relations? A Multidimensional Approach to Intersemiotic Convergence in Multimodal Texts," *Cognitive Linguistics* 4 (32), 529–562.

Haspelmath, Martin. 2023 "On What a Construction Is," *Constructions* 15(1).

Hilpert, Martin. 2014. *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.

Hirrel, Laura. 2018. "Cyclic Gestures and Multimodal Symbolic Assemblies: An Argument for Symbolic Complexity in Gesture," https://digitalrepository.unm.edu/ling_etds/57.

Hoffmann, Thomas. 2016. "Multimodal Constructs – Multimodal Constructions? The Role of Constructions in the Working Memory," *Linguistics Vanguard* 3.

Hoffmann, Thomas. 2021. "Multimodal Construction Grammar: From Multimodal Constructs to Multimodal Constructions," in Xu Wen and John R. Taylor (eds.), *The Routledge Handbook of Cognitive Linguistics*, 78–92.

Hoffmann, Thomas and Alexander Bergs. 2018. "A Construction Grammar Approach to Genre," *CogniTextes* 18, 1–27.

Hoffmann, Thomas, and Graeme Trousdale. 2013. *The Oxford Handbook of Construction Grammar*. New York: Oxford University Press.

Holmes, Jasper and Richard A. Hudson. 2005. “Constructions in Word Grammar,” in Jan-Ola Östman and Mirjam Fried (eds.), *Construction Grammars: Cognitive grounding and theoretical extensions* 3, 243–272.

House, Juliane. 2003. “English as a Lingua Franca: A Threat to Multilingualism?,” *Journal of Sociolinguistics* 7.4, 556–578.

Höder, Steffen. 2021. “Multilingualism and Diasystematic Construction Grammar,” *Diadorim* (1)23, 34–43.

Hyland, Ken. 2015. *Academic Publishing: Issues and Challenges in the Construction of Knowledge*. Oxford: Oxford University Press.

Hu, Yuxiu. 2016. “A Longitudinal Study on the Extent of Mandarin Influence on the Acquisition of English,” *International Journal of Language Studies* 10(4), 59–76.

Hyon, Sunny. 2017. *Introducing genre in English for Specific Purposes*. London: Routledge.

Ibbotson, Paul. 2013. “The Scope of Usage-based Theory,” *Frontiers in Psychology* 4.

Ingvarsdóttir, Hafdis, and Birna Arnbjörnsdóttir. 2013. “ELF and Academic Writing: A Perspective from the Expanding Circle,” *Journal of English as a Lingua Franca* 2(1), 123–145.

Jenkins, Jennifer. 2006/2009. “English as a Lingua Franca: Interpretations and Attitudes,” *World Englishes* 28, 200–207.

Jenkins, Jennifer. 2013. *English as a Lingua Franca in the International University: The Politics of Academic English Language Policy*. Abingdon: Routledge.

Jenkins, Jennifer. 2015. “Repositioning English and Multilingualism in English as a Lingua Franca,” *Englishes in Practice* 2(3), 49–85.

Jenkins, Jennifer. 2017. “ELF and WE: Competing or Complementing Paradigms?,” In Low Ee Ling and Anne Pakir (eds.), *World Englishes: Re-thinking Paradigms*.

Jenkins, Jennifer, Alessia Cogo, and Martin Dewey. 2010. “Review of Developments in Research into English as a Lingua Franca,” *Language Teaching* 44(3), 281–315.

Johansson, Robert and Rolf Ulrich. 2024. “Serial Processing of Proximity Groups and Similarity Groups,” *Attention, Perception, & Psychophysics* 86. https://doi.org/10.3758/s13414-024-02861-2

Kardela, Henryk. 2012. *Analyzability and Institutionalization: Setting Parameters in Cognitive Morphology*. The Free Library 01.

Kecskés, Istvan. 2014. *Intercultural pragmatics*. Oxford: Oxford University Press.

Kecskés, Istvan. 2019. *English as a Lingua Franca: The Pragmatic Perspective*. Cambridge: Cambridge University Press.

Kegalj, Jana, and Mirjana Borucinsky. 2021. "Genre-based Approach to Corpus Compilation for Translation Research," *7th International e-Conference on Studies in Humanities and Social Sciences*, 215–224.

Kiefer, Markus, and Lawrence Barsalou. 2013. "Grounding the Human Conceptual System in Perception, Action, and Internal States," In Wolfgang Prinz, Miriam Beisert, and Arvid Herwig (eds.), *Action Science: Foundations of an Emerging Discipline*, 381–407.

Kimura, Daisuke, and Suresh Canagarajah. 2017. "Translingual practice and ELF," In Jennifer Jenkins, Will Baker, Martin Dewey (eds.), *The Routledge Handbook of English as a Lingua Franca*, 295–308.

Kövecses, Zoltan. 2015. *Where Metaphors Come From: Reconsidering Context in Metaphor*. Oxford: Oxford University Press.

Kuo, I.-Chun. 2006. "Addressing the Issue of Teaching English as a Lingua Franca," *ELT Journal* 60 (3), 213–221.

Kuteeva, Maria. 2023. *Tension-Filled English at the Multilingual University: A Bakhtinian Perspective*. Bristol: Multilingual Matters.

Laitinen, Mikko. 2018. "What Happens to Ongoing Change in Multilingual Settings? A Corpus Compiler's Perspective on New Data and New Research Prospects," *Challenging the Myth of Monolingual Corpora* 80, 58–79.

Laitinen, Mikko, Magnus Levin, and Alexander Lakaw. 2019. "Charting New Sources of elf Data: A Multi-Genre Corpus Approach," *From Data to Evidence in English Language Research* 83, 326–350.

Lakoff, George. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press.

Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, George and Mark Johnson. 1999. *Philosophy in the Flesh: The Embodied Mind and Its Challenge to Western Thought*. New York: Basic Books.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar*. Stanford: Stanford University Press.

Langacker, Ronald W. 1999. "Correspondences, Compositionality, and Grammar," *MANUSYA Journal of Humanities* 2(2), 61–76.

Langacker, Ronald W. 2001. “Discourse in Cognitive Grammar,” *Cognitive Linguistics* 12(2), 143–188.

Langacker, Ronald W. 2005. “Construction Grammars: cognitive, radical, and less so,” In Francisco J. Ruiz de Mendoza Ibáñez and M. Sandra Peña Cervel (eds.), *Cognitive Linguistics: Internal Dynamics and Interdisciplinary Interaction*, 101–162.

Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Langacker, Ronald W. 2009. Investigations in Cognitive Grammar. Berlin: Mouton de Gruyter.

Langacker, Ronald W. 1993. “Reference-Point Constructions,” *Cognitive Linguistics* 4, 1–38.

Langacker, Ronald W. 2014. “Culture and Cognition, Lexicon and Grammar,” In Masataka Yamaguchi et. al. (eds.), *Approaches to Language, Culture and Cognition*, 27–49.

Langacker, Ronald W. 2016a. “Working Toward a Synthesis,” *Cognitive Linguistics*, 27(4), 465–477.

Langacker, Ronald W. 2016b. “Cognitive Grammar,” In Barbara Dancygier (ed.), *The Cambridge Handbook of Cognitive Linguistics*, 262–283.

Langacker, Ronald W. 2017a. “Evidentiality in Cognitive Grammar,” In J. I. Marín Arrese, G. Haßler, & M. Carretero (eds.), *Evidentiality Revisited*, 13–55.

Langacker, Ronald W. 2017b. *Ten Lectures on the Elaboration of Cogninitve Grammar*. Leiden/Boston: Brill.

Langacker, Ronald W. 2017c. “Entrenchment in Cognitive Grammar,” In Hans-Jörg Schmid (ed.), *Entrenchment and the Psychology of Language Learning: How we Reorganize and Adapt Linguistic Knowledge*, 39–56.

Langlotz, Andreas. 2010. “Social Cognition,” in Miriam A. Locher and Sage L. Graham (eds.), *Interpersonal Pragmatics*, 167–204.

Langlotz, Andreas. 2015. “Local Meaning-Negotiation, Activity Types, and the Current-Discourse-Space Model,” *Language and Cognition* 7.4, 515–545.

Larson-Hall, Jenifer. 2016. *A Guide to Doing Research in Second Language Acquisition with SPSS and R*. New York: Routledge.

van Leeuwen, Cees and Andrey Nolaev. 2011. “Gestalt Has No Notion of Attention. But Does It Need One?,” *Humana Mente. Journal of Philosophical Studies*, 35–68.

Li, Yongyan, and John Flowerdew. 2007. “Shaping Chinese Novice Scientists' Manuscripts for Publication,” *Journal of Second Language Writing* 16(2), 100–117.

Lidwell, William, and Kritina Holden. 2010. “Universal Principles of Design,” *Handbook of Applied Linguistics*.

Lindstedt, Jouko. 2009. “Esperanto: An East European Contact Language?,” In Christian Voß and Alicja Nagórko (eds.), *Die Europäizität der Slawia oder dies Slawizität Europas. Ein Beitrag der Kultur- und Sprachrelativistischen Linguistik*, 125–134.

Lorés-Sanz, Rosa. 2016. “ELF in the Making? Simplification and Hybridity in Abstract Writing,” *Journal of English as a Lingua Franca* 1(5), 53–81.

Löbner, Sebastian. 2002. *Understanding Semantics*. London: Hodder Arnold.

Löbner, Sebastian. 2015. “Functional Concepts and Frames,” In Thomas Gamerschlag, Doris Gerland, Rainer Osswald and Wiebke Petersen (eds.), *Meaning, frames, and conceptual representation*, 15–42.

Lüdeling, Anke and Kytö Merja. 2008. *Corpus Linguistics: An International Handbook*. Berlin/New York: Mouton de Gruyter.

Mackenzie, Ian. 2014. *English as a Lingua Franca: Theorizing and Teaching English*. London / New York: Routledge.

Mackenzie, Ian. 2015. “Will English as a Lingua Franca Impact on Native English?,” In Christina Sanchez-Stockhammer (ed.), *Can We Predict Linguistic Change?*

Makoni, Sinfree and Alastair Pennycook. 2012. “Disinventing Multilingualism: From Monological Multilingualism to Multilingual Francas,” in Marilyn Martin-Jones, Adrian Blackledge, and Angela Creese (eds.), *The Routledge Handbook of Multilingualism*, 439–453.

Mandler, Jean M. and Cristóbal Pagán Cánovas. 2014. “On Defining Image Schemas,” *Language and Cognition* 0, 1–23.

Mason, Oliver. 2008. “Stringing Together a Sentence. Linearity and the Lexis-Syntax Interface,” in Andrea Gerbig and Oliver Mason (eds.), *Language, People, Numbers: Corpus Linguistics and Society*, 231–248.

Matras, Yaron. 2009. *Language Contact*. Cambridge: Cambridge University Press.

Mauranen, Anna. 2009. “Chunking in ELF: Expressions for Managing Interaction,” *Intercultural Pragmatics* 6(2), 217–233.

Mauranen, Anna. 2006. “A Rich Domain of ELF – the ELFA Corpus of Academic Discourse,” *Nordic Journal of English Studies* 5.2, 145–159.

Mauranen, Anna. 2010. “Features of English as a Lingua Franca in Academia,” *Helsinki English Studies* 6, 6–28.

Mauranen, Anna. 2012. *Exploring ELF. Academic English Shaped by Non-Native Speakers*. Cambridge: Cambridge University Press.

Mauranen, Anna. 2013. “Lingua Franca Discourse in Academic Contexts: Shaped by Complexity,” In John Flowerdew (ed.), *Discourse in Context: Contemporary Applied Linguistics* 3, 225–246.

Mauranen, Anna. 2015. “English as a Global Lingua Franca: Changing Language in Changing Global Academia,” In Kumiko Murata (ed.), *Exploring ELF in Japanese Academic and Business Contexts: Conceptualization, Research and Pedagogic Implications*, 29–46.

Mauranen, Anna. 2016. “Temporality in speech – Linear Unit Grammar,” In *English Text Construction.*

Mauranen, Anna. 2017. “Conceptualising ELF.” In Jennifer Jenkins, Will Baker, and Martin Dewey (eds.), *The Routledge Handbook of English as a Lingua Franca,* 7–24.

Mauranen, Anna. 2018. “Linear Unit Grammar”, In Jan-Ola Östman and Jef Verschueren (eds.), *Handbook of Pragmatics*, 25–48.

Mauranen, Anna. 2019. “Linear Unit Grammar,” In J. Verschueren and J-O. Östman (eds.) *Handbook of Pragmatics* 21, 25–48.

Mauranen, Anna. 2021. “Good Texts in Non-Standard English,” In Kumiko Murata (ed.), *ELF Research Methods and Approaches to Data and Analyses*, 57–80.

Mauranen, Anna. 2023. *Reflexively Speaking: Metadiscourse in English as a Lingua Franca.* Berlin: de Gruyter.

Mauranen, Anna, and John McHardy Sinclair 2006. *Linear Unit Grammar: Integrating Speech and Writing*. Amsterdam / Philadelphia: John Benjamins.

Mauranen, Anna, Niina Hynnien, and Elina Ranta. 2016. “English as the Academic Lingua Franca,” *English for Specific Purposes* 29.3, 183–190.

McIntyre, Dan. 2020. *History of English: A Resource Book for Students*. New York: Routledge.

Meierkord, Christiane. 2004. “Lingua Franca Communication in Multiethnic Context,” In Helga Kotthoff and Helen Spencer-Oatey (eds.), *Handbook of Intercultural Communication*, 199–218.

Meierkord, Christiane. 2006. “Lingua Francas as Second Languages,” In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 163–171.

Meierkord, Christiane. 2007. “Lingua Franca Communication in Multiethnic Contexts,” In Helga Kotthoff, and Helen Spencer-Oatey (eds.), *Handbook of Intercultural Communication*, 199–218.

Mesthrie, Rajend and Rakesh Bhatt. 2008. “World Englishes: The study of new linguistic varieties,” *Key Topics in Sociolinguistics*.

Mesthrie, Rajend. 2009. “Pidgin/Creoles and Contact Languages: An Overview,” In Silvia Kouwenberg and John Victor Singler (eds.), *The Handbook of Pidgin and Creole Studies*, 263–286.

Millot, Philippe. 2015. “Defining English as a Professional Lingua Franca: A Specialised Approach,” *ASp* 67, 5–26.

Milroy, Lesley. 2002. “Social Networks,” In Jack K. Chambers, Peter Trudgill, and Natalie Schilling-Estes (eds.), *The Handbook of Language Variation and Change*, 549–572.

Nasar, Zara, Syed Waqar Jaffry, and Muhammad Kamran Malik. “Information Extraction from Scientific Articles: a Survey,” Scientometrics, 1-60.

Nielsen Norman Group. 2016. http://www.nngroup.com/articles/f-shaped-pattern-reading-web- content.

Nuttall, Louise. 2018. *Mind Style and Cognitive Grammar: Language and Worldview in Speculative Fiction*. London: Bloomsbury.

Operstein, Natalie. 2022. *The Lingua Franca. Contact-Induced Language Change in the Mediterranean*. Cambridge: Cambridge University Press.

O’Regan, John P. 2014. “English as a Lingua Franca: An Immanent Critique”, *Applied Linguistics* 35(5), 533–552.

Overmann, Karenleigh A. and Lambros Malafouris. 2018. “Situated cognition,” *International Encyclopedia of Anthropology: Anthropology Beyond Text*, 1–8.

Östman, Jan-Ola. 1999. “Coherence Through Understanding Through Discourse Patterns,” In Wolfram Bublitz, Uta Lenk, Eija Ventola (eds.), *Coherence in Spoken and Written Discourse*, 77–100.

Östman, Jan-Ola. 2005. “Construction Discourse: A Prolegomenon,” In Jan-Ola Östman, Mirjam Fried (eds.) Construction Grammars: *Cognitive Grounding and Theoretical Extensions, 121-140.*

Östman, Jan-Ola, and Mirjam Fried (eds.). 2005. *Construction Grammars: Cognitive grounding and theoretical extensions*. Amsterdam / Philadelphia: John Benjamins.

Page, Ruth. 2009. *New Perspectives on Narrative and Multimodality*. New York / London: Routledge.

Palmer, Stephen E. 2003. “Perceptual Organization in Vision,” In Steven Yantis (ed.), *Stevens' Handbook of Experimental Psychology. Volume 1: Sensation and Perception*, 177–234.

Patt, Sebastian. 2013. Punctuation as a Means of Medium-Dependent Presentation Structure in English. Tübingen: Verlag.

Paradis, Carita. 2001. “Adjectives and Boundedness,” *Cognitive Linguistics* 12(1), 47–65.

Paradis, Carita. 2005. “Ontologies and Construals in Lexical Semantics,” *Axiomathes* 15, 541–573.

Paradowski, Michał B. 2013. “Barbara Seidlhofer: Understanding English as a Lingua Franca: A Complete Introduction to the Theoretical Nature and Practical Implications of English used as a Lingua Franca (Review),” *The Interpreter and Translator Trainer* 7(2), 312–320.

Pecher, Diane and Rolf. A. Zwaan. 2005. *Grounding Cognition: The Role of Perception and Action in Memory, Language and Thinking*. Cambridge: Cambridge University Press.

Phillipson, Robert. 2010. *Linguistic Imperialism Continued*. London / New York: Routledge.

Pinker, Steven. 1984. *Language Learnability and Language Development*. Cambridge, MA: Harvard University Press.

Pinker, Steven. 2003. *The Language Instinct: How the Mind Creates Language*. UK: Penguin.

Pinna, Baingio. 2012. “Perceptual Organization of Shape, Color, Shade, and Lighting in Visual and Pictorial Objects,” *Iperception* 3(5), 257–281.

Pinna, Baingio and Katia Deiana. 2015. “The Syntax Organization of Shape and Color and the Laws of Coloration in Vision, Art and Biology,” *Art & Perception* 3, 319–345.

Pirc, Ana. 2013. “The Constructionist Modelling of ‘Non-native Discourse’,” *16èmes Rencontres Jeunes Chercheurs*, 80–92.

Pitzl, Marie-Luise. 2016. “World Englishes and Creative Idioms in English as a Lingua Franca,” *World Englishes* 35 (2), 293–309.

Pitzl, Marie-Luise. 2018. *Creativity in English as a Lingua Franca: Idiom and Metaphor*. Berlin/Boston: De Gruyter Mouton.

Pontello, Matteo. 2020. *English as a Lingua Franca and the Implications of ELF for English Teachers in Italy*. Venice: Ca’ Foscari University of Venice.

Prodromou, Luke. 2007. “A Reader Responds to J. Jenkins’s ‘Current Perspectives on Teaching World Englishes and English as a Lingua Franca,’” *Tesol Quarterly* 41.2. 409–413.

Quinn, Paul, Ramesh Bhatt, and Angela Hayden. 2008. “Young Infants Readily Use Proximity to Organize Visual Pattern Information,” *Acta Psychologica* 127(2), 289–298.

Radden, Günter, and René Dirven. 2007. *Cognitive English Grammar*. Amsterdam / Philadelphia: John Benjamins Publishing Company.

Radden, Günter and Zoltan Kövecses. 2007. “Towards a Theory of Metonymy,” In Klaus-Uwe Panther and Günter Radden (eds.), *Metonymy in Language and Thought*, 17–59.

Ranta, Elina. 2009. “Syntactic Features in Spoken ELF-Learner Language or Spoken Grammar,” In Anna Mauranen and Elina Ranta (eds.), *English as a Lingua Franca: Studies and Findings*, 84–106.

Ranta, Elina. 2013. “Universals in a Universal Language? Exploring Verb-Syntactic Features in English as a Lingua Franca.” PhD Diss., University of Tampere.

Ranta, Elina. 2017. “Grammar in ELF,” In Jennifer Jenkins, Will Baker, and Martin Dewey (eds.), *The Routledge Handbook of English as a Lingua Franca*, 244–254.

Rausch, Edwin. 1966. “The Problem of Properties in the Gestalt Theory of Perception,” In Wolfgang Metzger and Heiner Erke (eds.), *Handbook of Psychology: Vol. 1 Perception and Consciousness*, 866–953.

Rock, Irvin and Stephen E. Palmer. 1990. “The Legacy of Gestalt Psychology,” *Scientific American* 263 (6), 84–90.

Romaine, Suzanne. 2017. *Pidgin and Creole Languages*. London: Routledge.

Römer, Ute and Ann Arbor. 2009. “English in Academia: Does Nativeness Matter?,” *Anglistik: International Journal of English Studies* 20 (2), 89–100.

Rowley-Jolivet, Elizabeth. 2017. “English as a Lingua Franca in Research Articles: The SciELF Corpus,” *ASp* 71, 145–158.

Sakoda, Kent, and Jeff Siegel. 2003. *Pidgin Grammar: An Introduction to the Creole language of Hawai’i*. Honolulu: Bess Press.

Salverda, Reinier. 2018. “Empires and their Languages: Reflections on the History and the Linguistics of Lingua Franca and Lingua Sacra,” In Ian T. Baldwin, Gerd Graßhoff, Jürgen Renn, Dagmar Schäfer, Robert Schlögl, Bernard F. Schutz (eds.), *Multilingualism, Lingua Franca and Lingua Sacra*, 13–79.

Samarin, William J. 1978. “Lingua Francas of the World,” In Joshua A. Fishman (ed.), *Readings in the Sociology of Language*, 660–672.

Saraceni, Mario. 2015. *World Englishes: A Critical Analysis*. London: Bloomsbury.

de Saussure, Ferdinand. [1916] 1983. *Course in General Linguistics.* Roy Harris (trans.). London: Duckworth.

Schmid, Hans-Jörg. 2003. “An Outline of the Role of Context in Comprehension." In: Ewald Mengel, Hans-Jörg Schmid und Michael Steppat (eds.), *Anglistentag 2002 Bayreuth. Proceedings,* 435–445.

Schmid, Hans-Jörg. 2016. “Why Cognitive Linguistics Must Embrace the Social and Pragmatic Dimensions of Language and How it Could Do so More Seriously,” *Cognitive Linguistics* 27(4), 543–557.

Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press.

Schneider, Edgar W. 2007. *Postcolonial English: Varieties around the World*. Cambridge: Cambridge University Press.

Seidlhofer, Barbara. 2004. “Research Perspectives on Teaching English as a Lingua Franca,” *Annual Review of Applied Linguistics* 24, 209–239.

Seidlhofer, Barbara. 2011. *Understanding English as a Lingua Franca*. Oxford: Oxford University Press.

Seidlhofer, Barbara. 2017. “Standard English and the Dynamics of ELF Variation” In Jennifer Jenkins, Will Baker, and Martin Dewey (eds.), *The Routledge Handbook of English as a Lingua Franca*, 85–100.

Sebba, Mark. 1997. *Contact Languages: Pidgins and Creoles*. London: Bloomsbury Publishing.

Seid, Samson. 2015. “Ethnic Language Shift Among the Nao People of Ethiopia,” In Elizabeth C. Zsiga, One Tlale Boyer, Ruth Kramer (eds.), *Languages in Africa: Multilingualism, Language Policy, and Education*, 102–109.

Seizov, Ognyan. 2014. *Political Communication Online: Structures, Functions, and Challenges*. New York: Routledge.

Seuren, Pieter. 2004. *Chomsky’s Minimalism*. Oxford: Oxford University Press.

Seymore, Kristie, Andrew McCallum, and Ronald Rosenfeld. 1999. “Learning Hidden Markov Model Structure for Information Extraction” In Proceedings of the AAAI’99 workshop machine learning for information extraction , 37–42.

Shapiro, Lawrence and Shannon Spaulding. 2021. “Embodied Cognition,” *The Stanford Encyclopedia of Philosophy Summer 2024 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <https://plato.stanford.edu/archives/sum2024/entries/embodied-cognition/>*.

Shapiro, Lawrence and Shannon Spaulding (eds.). 2024. *The Routledge Handbook of Embodied Cognition* (2nd ed.). London / New York: Routledge.

Siegel, Jeff. 1990. “Pidgin English in Nauru,” *Journal of Pidgin and Creole Languages* 5(2), 157–186.

Sīle, Linda, Raf Guns, Frédéric Vandermoere, Gunnar Sivertsen, and Tim C. E. Engels. 2021. “Tracing the Context in Disciplinary Classifications: A Bibliometric Pairwise Comparison of Five Classifications of Journals in the Social Sciences and Humanities,” *Quantitative Science Studies* 2(1), 65–88.

Sinclair, John. 1996. “The Search for Units of Meaning.” *Textus* 9(1), 75–106.

Smith, S. M. 1988. “Environmental context-dependent memory,” In Graham M. Davis and Donald M. Thomson (eds.), *Memory in Context: Context in Memory*, 13–33.

Smith, Geoff P. and Jeff Siegel. 2013. “Tok Pisin,” *The Survey of Pidgin and Creole Languages, v. I: English-based and Dutch-based Languages*, 214–222.

Sobkowiak, Włodzimierz. 2005. “Why not LFC?,” In Katarzyna Dziubalska-Kołaczyk and Joanna Przedlacka (eds.), *English Pronunciation Models: A Changing Scene*, 131–149.

Solomon, Karen, and Lawrence Barsalou 2001. “Representing Properties Locally,” *Cognitive Psychology* 43(2), 129–169.

Spicer, Scott. 2014. “Exploring Video Abstracts in Science Journals: An Overview and Case Study,” *Journal of Librarianship and Scholarly Communication* 2(2), eP1110.

Spillmann, Lothar. 2009. “Phenomenology and Neurophysiological Correlations: Two Approaches to Perception Research,” *Vision Research* 49, 1507–1521.

Stadler, Michael W. 2020. *The Ontological Nature of Part-Whole Oscillations*. Wien: Verlag der Österreichischen Akademie der Wissenschaften.

Stefanowitsch, Anatol and Susanne Flach. 2017. “The Corpus-Based Perspective on Entrenchment,” In Hans-Jörg Schmid (ed.), *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*, 101–128.

Strugielska, Ariadna and Dorota Guttfeld. 2024. “Hyperlinks in Online Abstracts of Research Articles: Disciplines in the 5.0 Academic Community,” in: *Language, Technology, and Humanities in Society 5.0 – an interdisciplinary perspective*. Göttingen: Vandenhoeck & Ruprecht.

Strugielska, Ariadna and Katarzyna Piątkowska. 2017, “A Plea For A Socio-cognitive Perspective On The Language-Culture-Cognition Nexus In Educational Approaches To Intercultural Communicative Competence,” *Review of Cognitive Linguistics* 15(1), 224–252.

Stöckl, Hartmut. 2004. “Typography: Body and Dress of a Text - a Signing Mode between Language and Image,” *Visual Communication* 4, 205–214.

Sung, Chit C., M. 2014. “English as a Lingua Franca and Global Identities: Perspectives from four Second Language Learners of English in Hong Kong,” *Linguistics and Education* 26, 31–39.

Swales, John. 1990. *Genre Analysis: English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Szawerna, Michał. 2012. “A Frame-theoretic Characterization of Comics,” *Acta Universitatis Wratislaviensis. Anglica Wratislaviensia* 50, 197–211.

Talmy, Leonard. 2000. *Toward a Cognitive Semantics, Vol. 1. Concept Structuring Systems*. Cambridge: The MIT Press.

Taylor, John R. 1995. *Linguistic Categorization. Prototypes in Linguistic Theory*. Oxford: Clarendon Press.

Timmis, Ivor. 2002. "Native-Speaker Norms and International English: A Classroom View", *ELT Journal* 3(56), 240–249.

Todd, Loreto. 2005. *Pidgins and Creoles*. London / New York: Routledge.

Todorovic, Dejan. 2008. "Gestalt Principles," *Scholarpedia* 3(12): 5345.

Todorovic, Dejan. 2011. "What is the Origin of the Gestalt Principles?" *Humana Mente* 4(17), 1–20.

Tomasello, Michael. 2003. *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge: Harvard University Press.

Trask, Robert. 1999. *Key Concepts in Language and Linguistics*. London: Routledge.

Trudgill, Peter. 2005. "Native-Speaker Segmental Phonological Models and the English Lingua Franca Core", In Katarzyna Dziubalska-Kołaczyk and Joanna Przedlacka (eds.), *English Pronunciation Models: A Changing Scene*, 77–98.

Trudgill, Peter. 2016. *Dialect Matters: Respecting Vernacular Language*. Cambridge: Cambridge University Press.

Trudgill, Peter, and Jean Hannah. 2017. *International English: A Guide to Varieties of English Around the World*. London: Routledge.

Truong, Dothang. 2024. *Data Science and Machine Learning for Non-Programmers: Using SAS Enterprise Miner*. Oxon: CRC Press.

Tryon, Darrell T, and Jean-Michel Charpentier. 2004. *Pacific Pidgins and Creoles: Origins, Growth, and Development*. Berlin/New York: Mouton de Gruyter.

Tummers, Jose, Kris Heylen and Dirk Geeraerts. 2005. "Usage-based Approaches in Cognitive Linguistics: A Technical State of the Art," *Corpus Linguistics and Linguistic Theory* 2(1), 225–261.

Tweedie, M. Gregory, and Robert C. Johnson. 2022. *Medical English as a Lingua Franca.* Berlin/Boston: de Gruyter.

Tversky, Barbara. 2001. "Spatial Schemas in Depictions," In Merideth Gattis (ed.), *Spatial Schemas and Abstract Thought*, 79–112.

Tyler, Andrea 2010. "Usage-Based Approaches to Language and Their Applications to Second Language Learning," *Annual Review of Applied Linguistics* 30, 270–291.

Ungerer, Tobias. 2023. “A Gradient Notion of Constructionhood,” *Constructions* 15, 1–20.

Ungerer, Tobiad and Stefan Hartmann. 2023. *Constructionist Approaches*. Cambridge: Cambridge University Press.

Vetchinnikova, Svetlana. 2015. “Usage-based Recycling or Creative Exploitation of the Shared Code?: The Case of Phraseological Patterning,” *Journal of English as a Lingua Franca* 4(2), 223–252.

Vetchinnikova, Svetlana, Alena Konina, Nitin Williams, Nina Mikusova and Anna Mauranen. 2022. “Perceptual Chunking of Spontaneous Speech: Validating a New Method with Non-Native Listeners,” *Research Methods in Applied Linguistics* 1(2), 1–5.

Vetchinnikova, Svetlana, Alena Konina, Nitin Williams, Nina Mikusova and Anna Mauranen. 2023. “Chunking up Speech in Real-Time: Linguistic Predictors and Cognitive Constraints,” *Language and Cognition* 15(3), 453–479.

Vikør, Lars. 2004. “Lingua Franca and International Language,” In Ulrich Ammon, Norbert Dittmar, Klaus J. Mattheier and Peter Trudgill (eds.), *Sociolinguistics. An International Handbook of the Science of Language and Society*, 328–335.

Vincente, Veruska S. 2007. “English-based Pidgins and Creoles: From Social to Cognitive Hypotheses of Acquisition,” *Revista Virtual de Estudos da Linguagem* 5.9, 1–30.

Wagemans, Johan, Jacob Feldman, Sergei Gapshtein and Ruth Kimchi. 2012. “A Century of Gestalt Psychology in Visual Perception: II. Conceptual and Theoretical Foundations,” *Psychological Bulletin* 138(6), 1218–1252.

Walczyński, Marcin. 2012. *A Living Language: Selected Aspects of Tok Pisin in the Press (on the Basis of Wantok Newspaper)*. Nysa: Publishing Office PWSZ.

Waller, Rob. 1980. “Graphic Aspects of Complex Texts: Typography as Macropunctuation,” In Paul Kolers, Merald Wrolstad, and Herman Bouma (eds.), *Processing of Visible Language*, 241–253.

Watkowska, Dorota. 2020. “Redundancy in ELF: A Corpus-based Study on Negative and Modal Concord,” *Anglica. An International Journal of English Studies* 30(2), 71−86.

Watkowska, Dorota. 2020. “A Corpus-based Study on Question Tags in ELF. An Attempt to Determine an Emergent Construction” *Anglica. An International Journal of English Studies* 29(2), 135−150.

Watts, Richard J. 2011. *Language Myths and the History of English*. New York: Oxford University Press.

Weinreich, Uriel. 1953. *Languages in Contact: Findings and Problems*. New York: Publications of the Linguistic Circle of New York.

Werlich, Egon. 1976. *A Text Grammar of English*. Germany: Quelle & Meyer.

Wertheimer, Max. 1923. "Laws of Organization in Perceptual Forms," *A Source Book of Gestalt Psychology*, 71−88.

West, Sarah. M. 2013. "A Neoliberal Love Spell: Race, Tourism and the Yucatan Peninsula, An Analysis of Televisa's Sortilegio," In Cameron McCarthy, Alicia Kozma, Karla Palma-Milano, Margaret Fitzpatrick, and Nicole Lamers (eds.), *Mobilized Identities: Mediated Subjectivity and Cultural Crisis in the Neoliberal Era*, 123–138.

Widdowson, Henry. 1997. "EIL, ESL, EFL: Global issues and local interests," *World Englishes* 16(1), 135−146.

Widdowson, Henry. 2000. "On the Limitations of Linguistics Applied," *Applied Linguistics* 21(1), 3–25.

Widdowson, Henry. 2016. "ELF, Adaptive Variability and Virtual Language," In Marie-Luise Pitzl and Ruth Osimk-Teasdale (eds.), *English as a Lingua Franca: Perspective and Prospects*, 31–38.

Widdowson, Henry. 2020. *On the Subject of English: The Linguistics of Language Use and Learning*. Berlin: de Gruyter.

Wiedemer, Marcos Luiz and Marcia dos Santos Machado Vieira. 2022. "Discursive Paradigm as (Proto)Construction: Linguistic Alternation Via Socio-Communicative Practices," *Variação em Português e em Outras Línguas Românicas*, 269–278.

Wilson, Margaret. 2002. "Six Views of Embodied Cognition," *Psychonomic Bulletin & Review* 9, 625–636.

Winford, Donald. 2006. "Reduced Syntax in (Prototypical) Pidgins," In Ljuljana Progovac, Kate Paesani, Eugenia Casielles, and Ellen Barton (eds.), *The Syntax of Non-Sententials*, 283–307.

Wray, Alison. 2002. "Formulaic Language in Computer-supported Communication: Theory Meets Reality," *Language Awareness* 11(2), 114–131.

Wu, Xue, Anna Mauranen, and Lei Lei. 2019. “Syntactic Complexity in English as a Lingua Franca Academic Writing,” *Journal of English for Academic Purposes* 43, 1–13.

Velupillai, Viveka. 2015. *Pidgin, Creoles & Mixed Languages: An Introduction*. Amsterdam / Philadelphia. John Benjamins Publishing Company.

Verhagen, Arie. 2010. “Construal aishernd Perspectivization,” In Dirk Geeraerts and Hubert Cuyckens (eds.), *The Oxford Handbook of Cognitive Linguistics*, 48–81.

Yamaguchi, Masataka, Dennis Tay, and Benjamin Blount. 2014. *Approaches to Language, Culture, and Cognition: The Intersection of Cognitive Linguistics and Linguistic Anthropology*. New York: Palgrave Macmillan.

Yilmaz, Selahattin. 2020. *A Constructional Analysis of Written Academic English as a Lingua Franca: The Case of Unedited and Edited Research Writing*. Georgia State University.

Yilmaz, Selahattin, and Ute Römer. 2020. “A Corpus-Based Exploration of Constructions in Written Academic English as a Lingua Franca,” In Ute Römer, Viviana Cortes, and Eric Friginal (eds.), *Advances in Corpus-based Research on Academic Writing: Effects of Discipline, Register, and Writer Expertise*, 59–87.

Yule, George. 2016. *The Study of Language*. Cambridge: Cambridge University Press.

Zenner, Eline and Dirk Geeraerts. 2018 “One Does Not Simply Process Memes: Image Macros as Multimodal Constructions,” In Esme Winter-Froemel and Verena Thaler (eds.), *Cultures and Traditions of Wordplay and Wordplay Research*, 167–194.

Żyśko, Konrad. 2017. “Analysability/Compositionality Parameter of Wordplay,” *Roczniki Humanistyczne* 11, 61–74.