

Univerzita Karlova
Pedagogická fakulta
Katedra anglického jazyka a literatury

DIPLOMOVÁ PRÁCE

Syntactic complexity of English L2 novice academic writing
Syntaktická komplexnost anglických odborných textů nerodilých mluvčích

Bc. Zuzana Ježková

Vedoucí práce: doc. PhDr. Markéta Malá, Ph.D

Studijní program: Učitelství anglického jazyka pro 2. stupeň základní školy a střední školy

Studijní obor: N AJ 20

Odevzdáním této diplomové práce na téma Syntaktická komplexnost anglických odborných textů nerodilých mluvčích potvrzuji, že jsem ji vypracovala pod vedením vedoucího práce samostatně za použití v práci uvedených pramenů a literatury. Dále potvrzuji, že tato práce nebyla využita k získání jiného nebo stejného titulu.

V Praze dne 8. 7. 2024

I would like to express my sincere gratitude to my thesis supervisor doc. PhDr. Markéta Malá, Ph.D. for her guidance, time, patience, and invaluable advice throughout the writing process.

ABSTRAKT

Diplomová práce se zabývá syntaktickou komplexností předmětů z kontrastivního lexikálně-gramatického úhlu pohledu. Jejím cílem je popsat a porovnat různé úrovně a aspekty strukturní komplexnosti předmětu v psaných odborně literárně-vědných textech tří typů autorů – odborníků a univerzitních studentů, jejichž mateřským jazykem je angličtina a univerzitních studentů angličtiny, jejichž mateřským jazykem je čeština. Materiál použitý v analytické části této práce byl čerpán ze tří korpusů současné psané akademické angličtiny – VESPA_CZ, BAWE a korpusu článků expertů publikovaných v akademických časopisech. Metodologie práce je založena na korpusově založeném přístupu a kontrastivní mezijazykové analýze (CIA). Teoretické části poskytují podrobný popis multidimenzionálního konstruktů komplexnosti a syntaktický popis předmětu, které jsou použity jako základ pro analýzu příkladů v analytické části této práce. Analytická část je založena na kombinaci kvantitativního a kvalitativního výzkumu a zkoumá různé parametry syntaktické komplexnosti na padesáti vzorcích z každé skupiny. Výsledky analýzy ukazují, že mezi sledovanými skupinami autorů existují významné rozdíly v úrovni komplexnosti realizace předmětu, které souvisejí jak se zkušeností autora s psáním odborného textu, tak s užíváním angličtiny jako mateřského nebo cizího jazyka. Diplomová práce také potvrdila, že komplexnost psaného akademického registru spočívá především v komplexní struktuře frází.

KLÍČOVÁ SLOVA

odborný psaný text, nerodilí studenti angličtiny, komplexnost, syntax, předmět, korpusově založený výzkum

ABSTRACT

The diploma thesis explores the syntactic complexity of objects from a contrastive lexicogrammatical point of view. It aims to describe and compare different levels and aspects of the structural complexity of the object in written academic texts from the field of literary history written by three types of authors: experts and university students with English as their first language, and university students of English with Czech as their first language. The material used in the analytical part of this thesis was drawn from three corpora of present-day written academic English – the VESPA_CZ, the BAWE and a corpus of articles published in academic journals. The methodology of the thesis rests on a corpus-assisted approach and the Contrastive Interlanguage Analysis (CIA). The theoretical parts give a detailed description of the multi-dimensional construct of complexity and a syntactic description of the object, both of which are used as the basis for the analysis of the examples in the analytical part of this thesis. The analytical part is based on a combination of quantitative and qualitative research and examines different parameters of syntactic complexity of 50 samples from each of the studied groups. The results show that there are significant differences in the level of complexity in the realisations of objects among the studied groups of authors. These differences are related both to the authors' experience with academic writing and to the use of English as a first or second language. The diploma thesis has also confirmed that the complexity of the written academic register lies primarily in the complex structure of phrases.

KEYWORDS

academic written text, English L2 students, complexity, syntax, object, corpus-assisted approach

Table of contents

List of abbreviations	8
1 Introduction	9
2 Theoretical background – linguistic complexity	11
2.1 Defining complexity	11
2.2 Types and components of complexity	14
2.3 Complexity features and devices	16
2.3.1 Structural type and syntactic function	16
2.3.2 Register differences	17
2.3.3 Hypothesised developmental stages for complexity features.....	17
2.4 Measuring complexity	19
2.4.1 Different measures of complexity	19
2.4.2 Computational measurement	20
3 Theoretical background – the object	22
3.1 Definition, form and position	22
3.2 Types of objects.....	23
3.2.1 Direct object	23
3.2.2 Indirect object.....	24
3.2.3 Prepositional object	25
3.3 Realisation forms of objects	26
3.4 Important related notions.....	27
3.4.1 Noun phrase (NP)	27
3.4.2 Determination, modification, complementation.....	27
3.4.3 Apposition	28
4 Material and method.....	30

4.1	Material.....	30
4.1.1	The Varieties of English for Specific Purposes dAtabase (VESPA_CZ) ...	30
4.1.2	British Academic Written English Corpus (BAWE).....	30
4.1.3	L1-expert Corpus.....	31
4.1.4	Register characteristics of written academic texts.....	31
4.1.5	Learner corpora.....	32
4.1.6	Academic and linguistic challenges	34
4.2	Method.....	34
4.2.1	Procedure.....	34
5	Analytical part	36
5.1	Main clause verbs	36
5.2	Transitivity, verb valency	38
5.3	Length of objects	40
5.4	Types of objects.....	41
5.4.1	Direct object	42
5.4.2	Indirect object.....	47
5.4.3	Prepositional object	47
5.5	Discussion in relation to developmental stages.....	48
5.5.1	Clausal realisations of objects	50
5.5.2	Phrasal realisations of objects – noun phrases.....	52
5.5.3	Phrasal realisations of objects – prepositional phrases.....	57
5.5.4	Coordination and apposition.....	59
6	Conclusions	61
	References	65
	Appendix 1 (L1-expert subcorpus: 50 examples analysed in the analytical part).....	68

Appendix 2 (L1-novice subcorpus: 50 examples analysed in the analytical part)	72
Appendix 3 (L2-novice subcorpus: 50 examples analysed in the analytical part)	75

List of abbreviations

A = adverbial

L1 = English as a first language

L2 = English as a second language

NP = noun phrase

O = object

O_d = direct object

O_i = indirect object

O_p = prepositional object

PP = prepositional phrase

S = subject

sb = somebody

SLA = second language acquisition

sth = something

V = verb

1 Introduction

This diploma thesis aims to describe and compare the different levels of syntactic complexity and specific types of complexity devices in structures that realise objects. The description and comparison are conducted within three separate groups of academic writers: L1 experts, L1 novices (VESPA_CZ), and L2 novices (BAWE). The object as the main point of interest in this thesis has been chosen as it can be expected that both clausal and phrasal realisations will occur, thus allowing for a thorough analysis of complexity and its devices. We approach complexity as a multi-dimensional construct which can be illustrated by the following quote.

“[...] [G]rammatical text complexity is a multi-dimensional linguistic construct, in the sense that groupings of phrasal/clausal complexity features co-occur in texts and pattern together as underlying parameters (or *dimensions*) of complexity, such that each text has a specific characterization in a multi-dimensional space of complexity.”
(Biber et al. 2023: 5)

The methodology adopted in this thesis is based on a lexico-grammatical approach, focusing not only on the types of objects but also on the controlling verbs and other associated patterns. Moreover, this thesis relies on a corpus-assisted approach performed within the Contrastive Interlanguage Analysis (CIA) framework, combining quantitative and qualitative research.

This diploma thesis is divided into four major parts: (1) the theoretical background of complexity; (2) the theoretical background of objects; (3) the material and method; and (4) the analytical part.

The aim of the theoretical part dealing with complexity is to define this construct, describe its types and components and explain different types of complexity features and devices. The emphasis is on the distinction between structural types and syntactic functions, register differences and the elucidation of hypothesised developmental stages for complexity features as proposed by Biber et al. (2011). Moreover, some types of measures used for measuring complexity are mentioned.

The theoretical part describing the object is designed to define this element and describe its typical forms and positions. In addition, all three types of objects are described in detail, and

their realisation forms are provided. Lastly, this part also gives a description of other related notions that are used in the analytical part.

The analytical part is subdivided into five parts, each focusing on a different parameter of the multi-dimensional construct of the structural complexity of objects in each of the three studied subcorpora. Namely, these parameters are the main clause verbs, the transitivity of these verbs, the length of the structures realising objects, different types of objects and their realisations, and the relation of the findings to the hypothesised developmental stages of complexity features. For the analysis, 50 samples from each of the subcorpus are used. The aim of the analytical part is to uncover the tendencies of writers from each of the groups and find similarities and differences.

The main research question of this thesis is how the complexity of objects in English academic written texts produced by Czech university students compares to similar texts written by English L1 university students and to professional published academic texts. Another research question is connected to the extent to which the differences can be attributed to English as a second language and English as the mother tongue contrast, and to what extent does the register of academic writing play a role.

2 Theoretical background – linguistic complexity

2.1 Defining complexity

When talking about linguistic complexity, it is necessary to clarify exactly what is meant by this term. The issue is that there is no widely acknowledged definition of complexity. In research, the term is either not defined at all or defined generally and vaguely, which results in ambiguity and confusion (Bulté and Housen 2012: 22, Biber et al. 2023: 2). Dictionary definitions of "complexity" such as "the state of being formed of many parts" or "the state of being difficult to understand" ("Complexity") illustrate the vagueness clearly as they are profoundly generic.

Many authors strictly divide the view on complexity into two major approaches – absolute and relative (Biber et al. 2023: 2). The absolute or objective approach defines complexity as "formal properties of the linguistic system" (Palotti 2014: 2) or as "the number of discrete components that a language feature or a language system consists of" and "the number of connections between the different components" (Bulté and Housen 2012: 24). Bulté and Housen (Ibid.) term complexity within this approach as "absolute complexity", "inherent complexity" or simply as "complexity", Palotti calls it "structural complexity" (2014: 2). In the relative or agent-related approach, the definition of complexity is that of "psychological difficulty in using or learning [linguistic units]" (Crystal 1997¹ in Biber et al. 2023: 2) or "the mental ease or difficulty with which linguistic items are learned, processed or verbalized" (Bulté and Housen 2012: 23). The terms used for complexity within this approach are "cognitive complexity" or "difficulty" (Ibid.). This thesis will be carried out within the absolute approach.

Additionally, Palotti proposes developmental or outsider complexity as an extra approach or meaning of complexity in linguistic research and defines it as "the order in which linguistic structures emerge and are mastered in second (and, possibly, first) language acquisition" (2014: 2).

¹ Crystal, David. *The Cambridge encyclopedia of language*. 2nd vol., Cambridge University Press, 1997.

As opposed to the straightforward division of absolute and relative complexity performed by many authors (e.g. Bulté and Housen 2012, Palotti 2014), Biber et al. (2023: 2) claim that it is also necessary to focus on the fact that the type of entity that is being characterized influences and causes differences in complexity. There are three types of entities relevant to linguistic research of complexity: (a) phrases/clauses (or sentences); (b) languages (or dialects); and (c) texts. Firstly, phrases/clauses are a subject of inquiry in descriptive grammars where the terms "simple" and "complex" stand at the forefront. Researchers adopting this approach examine the possibility of modifying phrases or clauses with additional optional elements. A phrase/clause which comprises only obligatory elements is regarded as simple, whereas the more optional elements a phrase/clause includes, the more complex it is considered. Dependent clauses have therefore been commonly seen as the primary indication of complexity within phrases and clauses (Biber et al. 2023: 2). What needs to be pointed out is that the traditional grammatical description of complexity does not usually rely on quantitative analysis, characterising instead the phrase or clause as complex if it comprises any optional elements (Ibid.: 3). Secondly, authors who investigate the complexity of languages or dialects go with the term "system complexity" as they focus on the linguistic system itself, hence using de Saussure's terminology, *langue* (Ibid.: 3; Palotti 2014: 4). This approach is adopted primarily by typologists and functional linguists, such as Nichols (2013²) or McWhorter (2001³). When the analysis carried out within this approach is quantitative, researchers use suitable variables (e.g. the number of grammatical distinctions made in the system) to calculate various features of the linguistic system (Biber et al. 2023: 3). Thirdly, for the purposes of this thesis most importantly, the focus may be on complexity in texts, which is often performed in studies within the Register-Functional (R-F) framework. In this approach, the notion of grammatical complexity is applied to actual language use, and this is accomplished by factoring register variation and lexico-grammatical patterns into analyses (Ibid.). A term used for this

² Nichols, Johanna. "The Vertical Archipelago: Adding the Third Dimension to Linguistic Geography." *Space in Language and Linguistics: Geographical, interactional, and cognitive perspectives*, edited by Peter Auer, Martin Hilpert, Anja Stukenbrock and Benedikt Szmrecsanyi, Walter de Gruyter, 2013.

³ McWhorter, John. "The World's Simplest Grammars Are Creole Grammars." *Linguistic Typology*, vol. 5, no. 2-3, 2001, pp. 125-166, *De Gruyter*, <https://doi.org/10.1515/lity.2001.001>.

understanding of complexity employed by Szmrecsanyi (2015⁴) is "usage-based or text complexity"; again, using de Saussure's terminology, this may be regarded as parole (Biber et al. 2023: 3; Palotti 2014: 4). Analyses of text complexity use measures aiming to depict the usage of different complexity features within a particular text. Therefore, these analyses may be labelled as mostly quantitative (Biber et al. 2023: 3).

Another distinction used while discussing complexity is whether it is employed as an independent or dependent variable in research. On the one hand, as an independent variable, complexity is an element that is being examined for its impact on a particular aspect of second language (L2) performance or proficiency. An example of employing complexity as an independent variable is the effect of instruction on second language acquisition (SLA) (Bulté and Housen 2012: 21). Palotti (2014: 1) states that when complexity is employed as an independent variable, it is used to refer to various features that either increase or decrease the complexity of a communicative task. On the other hand, as a dependent variable, complexity refers to diverse aspects of linguistic production (Ibid.), and when explored in this way, it is often accompanied by fluency and accuracy. In this view, complexity solely describes L2 performance and serves as an indicator of proficiency (Bulté and Housen 2012: 21).

Biber et al. (2023: 5) divide studies of complexity into two major types – those that regard complexity as a single and unified construct (e.g. Wolfe-Quintero et al. 1998⁵; Lu 2011⁶, 2017) and those that don't (Norris and Ortega 2009⁷; Ortega 2015⁸.; Bulté and Housen 2014⁹; Biber et al. 1992¹⁰, 2011, 2023). This thesis will join the latter approach, regarding

⁴ Szmrecsanyi, Benedikt. "Recontextualizing Language Complexity." *Change of Paradigms: New Paradoxes: Recontextualising Language and Linguistics. Applications of Cognitive Linguistics*, edited by Jocelyne Daems, Eline Zenner, Kris Heylen, Dirsk Speelman and Hubert Cuyckens, vol. 31, De Gruyter, 2015, pp. 347-360.

⁵ Wolfe-Quintero, Kate, et al. *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*, University of Hawaii, 1998.

⁶ Lu, Xiaofei. "A Corpus-based Evaluation of Syntactic Complexity in Second Language Writing." *TESOL Quarterly*, vol. 45, no. 1, 2011, pp. 36-61.

⁷ Norris, John M., and Lourdes Ortega. "Towards an Organic Approach to Investigating CAF in SLA: The Case of Complexity." *Applied Linguistics*, vol. 30, no. 4, 2009, pp. 555-578.

⁸ Ortega, Lourdes. "Syntactic Complexity in L2 Writing: Progress and Expansion." *Journal of Second Language Writing*, vol. 29, 2015, pp. 404-415.

⁹ Bulté, Bram, and Alex Housen. "Conceptualizing and Measuring Short-term Changes in L2 Writing Complexity." *Journal of Second Language Writing*, vol. 26, 2014, pp. 42-65.

¹⁰ Biber, Douglas. "On the Complexity of Discourse Complexity: A Multidimensional Analysis." *Discourse Processes*, vol 15, 1992, pp. 133-163.

complexity as a multi-dimensional construct. It is important to realise that all of the above-mentioned distinctions (e.g. text complexity vs. system complexity) do not represent different terms but different constructs. Even though different authors use various names for the same notion, it needs to be pointed out that these names serve as a simplification and the constructs behind them are what is really essential (Palotti 2014: 3). Biber et al. state that: "[G]roupings of phrasal/clausal complexity features co-occur in texts and pattern together as underlying parameters (or dimensions) of complexity, such that each text has a specific characterization in a multi-dimensional space of complexity." (2023: 5).

2.2 Types and components of complexity

As has already been pointed out, it is possible to analyse complexity from several points of view since it is a multi-dimensional construct. Thus, it is crucial to be knowledgeable in the types and components of L2 complexity. Bulté and Housen (2012: 23) offer a transparent taxonomy of the possible complexity constructs (cf. Figure 1). Some of the constructs in the taxonomy have already been described above (e.g. relative vs. absolute complexity), and the other constructs relevant to this thesis will be described below in more detail.

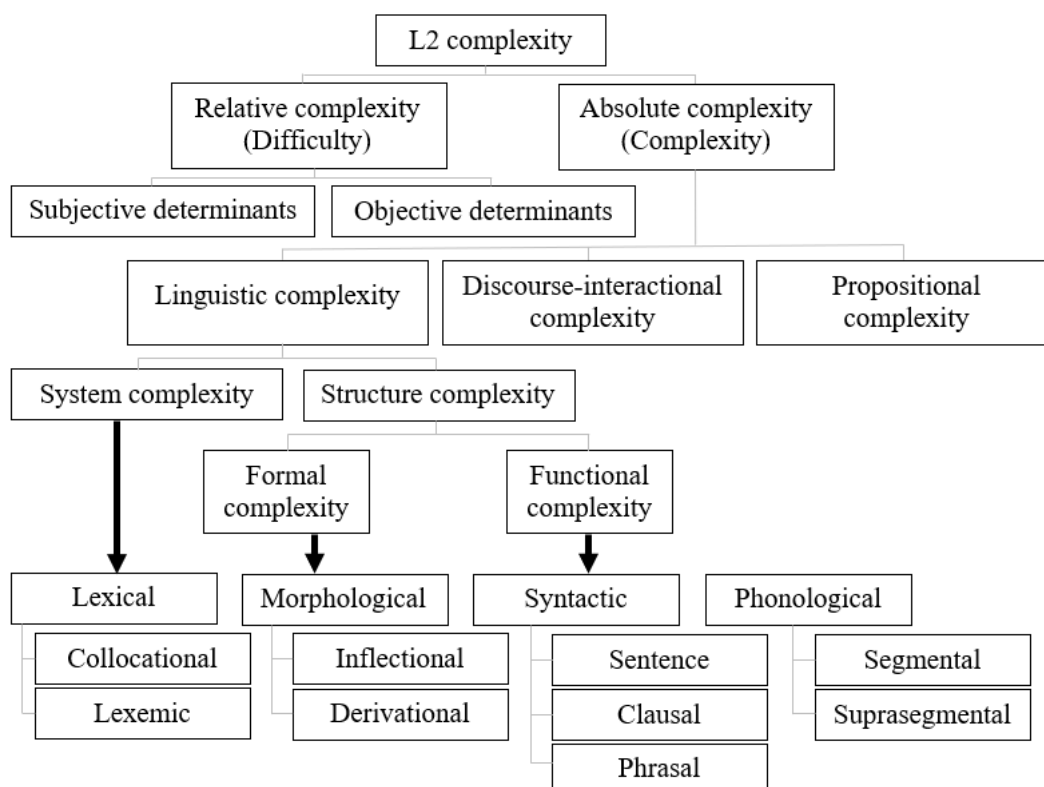


Figure 1: A taxonomy of complexity constructs (Bulté and Housen 2012: 23)

When examined from a broader point of view, complexity is argued to be composed of three major parts, namely, (a) propositional complexity; (b) discourse-interactional complexity; and (c) linguistic complexity (Bulté and Housen 2012: 24). Firstly, propositional complexity may be defined as "the number of information or idea units which a speaker/writer encodes in a given language task to convey a given message content" (Ibid.). Secondly, discourse-interactional complexity is a somewhat equivocal part of L2 complexity, primarily employed while investigating dialogues. The definition of this component is that of "the number and type of turn changes that learners initiate and the interactional moves and participation roles that they engage in" (Bulté and Housen 2012: 25). Thirdly, linguistic complexity can be further subdivided into (a) global/system complexity and (b) local/structure complexity. The main difference between the two subcomponents is that system complexity is more concerned with the range of particular words and structures, whereas structure complexity primarily focuses on the depth of relationships between different linguistic features, structures and rules (Ibid.). Moreover, there are two sub-branches of local complexity, namely, (a) functional and (b) formal complexity. Functional complexity can be described as "the number of meanings and functions of a linguistic structure and the degree of transparency, or multiplicity, of the mapping between the form and meanings/functions of a linguistic feature" (Ibid.). In contrast, defining formal complexity is not as straightforward as defining functional complexity, and since examining form will not be of concern in this thesis, the definitions do not need to be provided. As shown in Figure 1, both formal and functional complexity dimensions can be further subdivided into more domains: lexical, morphological, syntactic and phonological. Furthermore, all of these domains are associated with specific subdomains; for example, the domain of syntactic complexity has the subdomains of sentence, clausal and phrasal complexity (Bulté and Housen 2012: 26).

What needs to be kept in mind is that all of the above-described constructs are solely theoretical, and in practical terms, they usually do not exist on their own as they are often interwoven. Moreover, even though Bulté and Housen's taxonomy of complexity constructs (2012: 23) is very transparent, it is not a theory of complexity as no such theory has yet been created. The result of the non-existence of such a theory is that there is no agreement among different authors as to what features are regarded as structurally simple and complex. This is visible in the inconsistency of research results regarding the same feature performed

by various authors. Therefore, the dire need for such a theory is unquestionable (Bulté and Housen 2012: 26).

Palotti (2014: 4) describes an additional division of complexity of (a) grammatical and (b) stylistic complexity. Grammatical complexity, in this sense, is defined as "the complexity of linguistic rules that must be followed to produce grammatical sentences" (Ibid.), whereas stylistic complexity deals with individual stylistic variation.

2.3 Complexity features and devices

2.3.1 Structural type and syntactic function

In both studies, Biber et al. (2011: 22, 2023: 7) categorise features of structural complexity based on two parameters – structural type and syntactic function. Biber et al. (Ibid.) identify three structural types (or grammatical structures) and three syntactic functions that together form a 9-dimension model of text complexity. The three structural types are (a) finite dependent clauses; (b) nonfinite dependent clauses; and (c) dependent phrases (Biber et al. 2023: 7). And the three syntactic functions are (a) clause constituent; (b) noun phrase (NP) constituent; and (c) other-phrase constituent (Ibid.). This distinction will be used in the analytical part of this thesis. Moreover, Biber et al. (2023: 7) also mention the possibility of including the distinction of specific syntactic functions, such as modifiers and complements.

As already mentioned, the nine dimensions create a model of text complexity that relies both on structural types and syntactic functions. Biber et al. (2023) compare this model with other complexity models, which either treat complexity as a unified construct or restrict the examination of complexity to the structural types or to syntactic functions only. The results of Biber et al.'s analysis (2023) have shown that the model that takes both structural types and syntactic functions into consideration provides the best fit. What is more, according to the analysis, the importance of the distinction between syntactic functions is, in fact, more important than the distinction between structural types; however, considering both is the best way to research textual co-occurring patterns (Ibid.: 24). This particular result is especially relevant for this thesis as it is one of the reasons why object and its realisations is the focal

point of this thesis. On top of that, the results confirm that complexity is not a construct that is unified but a multi-dimensional construct (Biber et al. 2023: 19).

2.3.2 Register differences

First, it should be realised that "speech and writing are grammatically complex in dramatically different ways" (Biber et al. 2011: 10). There have been two major stereotypes regarding complexity in different registers. Firstly, clausal subordination is the most suitable tool for measuring complexity. Secondly, if clausal subordination is the best tool, then "academic writing is obviously more complex than conversation" (Biber et al. 2011: 10). However, the results of the research carried out by Biber et al. (2011) show that when treating complexity as a multi-dimensional construct, academic writing and conversation do not differ in the degree of complexity, but rather in the prevalent sources of complexity.

Biber et al. (2011: 22) performed a corpus analysis in which they employed the above-described distinction of structural type and syntactic function on the two registers, conversation and academic writing. The findings summarise that the complexity of conversation is clausal, whereas the complexity of academic writing is phrasal (Ibid.). Moreover, their research shows a significant difference in the types of complexity devices in the two registers, as "most of the complexity features are strongly favoured in either conversation or academic writing, not both" (Ibid.). On the whole, the findings show that it is irrelevant to consider clausal subordination as the only indicator of text complexity.

2.3.3 Hypothesised developmental stages for complexity features

On the basis of their research results, Biber et al. (2011) present a table of hypothesised developmental stages for complexity features (cf. Table 1), which was created by implications from the comparison of the two registers, conversation and academic writing. The stages are based on several assumptions described in the following lines. The first and most important assumption is that "[c]onversation is acquired first; the grammar of writing is acquired later, and not always successfully" (Biber et al. 2011: 29). The second assumption is connected to the question of difficulty. Biber et al. (Ibid.) state that those structures that are easily and early acquired, frequent in L1 conversations, "do not represent a high degree of production complexity" (Ibid.). In contrast, structures acquired later and with a relatively low frequency of use "represent a considerably higher degree of production complexity than

the conversational complexity features” (Ibid.). The third assumption revolves around the notion of lexical grammar. Biber et al. (2011: 30) place emphasis on the importance of lexicogrammatical factors and claim that the more fixed a lexicogrammatical combination is, the earlier it is acquired (Ibid.: 31). In connection with that, it is possible to analyse certain groupings of complexity features that often co-occur together, and these may be regarded as "underlying dimensions of grammatical text complexity" (Biber et al. 2023: 25). All in all, the developmental stages proposed by Biber et al. (2011) are a very suitable foundation for the assessment of student writing and in this thesis, they will be used to evaluate both L1 and L2 complexity in writing.

Stage	Grammatical structure(s)
1	Finite complement clauses controlled by common verbs (e.g. <i>think, know, say</i>)
2	Finite complement clauses controlled by a wider set of verbs
	Finite adverbial clauses
	Nonfinite complement clauses controlled by common verbs (esp. <i>want</i>)
	Phrasal embedding in the clause: adverbs and adverbials
	Simple phrasal embedding in the NP: attributive adjectives
3	Phrasal embedding in the clause: PPs as adverbials
	Finite complement clauses controlled by adjectives
	Nonfinite complement clauses controlled by a wider set of verbs
	<i>That</i> relative clauses, esp. with animate head nouns
	Simple phrasal embedding in the NP: nouns as pre/postmodifiers
	Possessive nouns as premodifiers
	<i>Of</i> phrases as postmodifiers
	Simple PPs as postmodifiers, esp. with prepositions other than <i>of</i> (concrete, locative meanings)
4	Nonfinite complement clauses controlled by adjectives
	Extraposed complement clauses
	Nonfinite relative clauses
	More phrasal embedding in the NP (attributive adjectives, nouns as premodifiers)
	Simple PPs as postmodifiers, esp. with prepositions other than <i>of</i> (abstract meanings)

5	Preposition + nonfinite complement clause
	Complement clauses controlled by nouns
	Appositive noun phrases
	Extensive phrasal embedding in the NP (multiple PPs as postmodifiers, with levels of embedding)

Table 1: Hypothesised developmental stages for complexity features (adapted from Biber et al. 2011: 30-31)

2.4 Measuring complexity

Generally speaking, measuring complexity is a very broad area which involves many issues. The studies focusing on measuring complexity can be divided into those that consider complexity as a unified construct and those that see it as a multi-dimensional construct. A vast majority of studies fall into the first group, relying on quantitative analysis with objective measures (Bulté and Housen 2012: 29). As already mentioned, the authors can also be divided into two groups according to which approach to complexity they choose. The problem with treating complexity as a multi-dimensional construct is that even those authors who attempt to treat it in this way often fail to do so as, in the end, they rely on omnibus measures (Biber et al. 2023: 5). Moreover, the area of complexity is composed of many sub-components; however, the number of measures designated to assess them is not identical for each dimension. In some cases, there are many measures for some sub-components (e.g. sentential syntactic complexity through subordination); in other cases, there are only one or two measures (e.g. lexical density) or even none at all (e.g. phrasal syntactic complexity) (Butlé & Housen 2012: 29). The issue with the range of measures could be easily avoided if, again, authors started treating complexity as a multi-dimensional construct, as Biber et al. (2011, 2023) propose and also support with strong empirical evidence. Therefore, the approach adopted in this thesis will follow Biber et al.'s (2011, 2023) view on complexity by treating it as a multi-dimensional construct and considering register characteristics.

2.4.1 Different measures of complexity

As already mentioned, there are many measures of complexity; some are suitable for measuring complexity while not treating it as a single unified construct, and some are not.

Moreover, as the number of complexity measures is rather high, the questions of reliability, validity, sensitivity, discriminatory power and practical feasibility arise and have to be dealt with (Bulté and Housen 2012: 35). Generally, the tendency in studies measuring complexity is to consider "one or more of the following: range of syntactic structure, length of unit, degree of structural complexity (sophistication) of certain syntactic structures, amount and type of coordination, subordination and embedding" (Ibid.). What is more, a significant number of measures can be labelled as ambiguous (or hybrid) as they capture multiple behavioural or theoretical complexity constructs and sources, which may be independent and unrelated at the same time (Ibid.). It also needs to be realised that each of the measures is somehow questionable in its computation or interpretation (Bulté and Housen 2012: 40).

The most commonly used measures are based on T-units¹¹ (e.g. mean length of T-unit, clauses per T-unit) (Bulté and Housen 2012: 30). However, according to Biber et al. (2011: 12), "there is little empirical evidence to recommend the use of T-unit-based measures for the study of grammatical complexity" because it has only limited use in testing applications, and it does not have any firm foundation in linguistic theory. In general, the measures can be divided according to the dimension of complexity they measure: (a) grammatical complexity or (b) lexical complexity, and then further subdivided according to the subdimensions (cf. 2.2) (Bulté and Housen 2012: 30-31). Bulté and Housen (Ibid.) provide a transparent table with forty measures of complexity in total, including the mean length of T-unit, dependent clauses per clause and mean length of clause for syntactic complexity; frequency of tensed forms and frequency of modals for morphological complexity. There are also measures of lexical complexity present in the table, but these do not need to be addressed here as lexical complexity will not be of concern in this thesis.

2.4.2 Computational measurement

As measuring complexity is very demanding, some authors tried to establish computational systems to measure and assess it. Using computational systems is a rather summarising measurement technique, but it will only be briefly described here as it does not fall into the approach adopted in this paper (i.e., treating complexity as a multi-dimensional construct).

¹¹ The author of the term T-unit is Hunt (1965: 37), who defines T-units as "the shortest grammatically allowable sentences into which the theme could be segmented".

The reason behind creating such systems was to make it easier to assess complexity using multiple measures (Lu 2010: 476). Some existing computational systems include Coh-Metrix, the Biber Tagger and the L2 Syntactic Complexity Analyzer (L2SCA) (Lu 2010, 2017). All of these systems have their advantages and disadvantages; Lu (2010, 2017) thoroughly describes the principle of his computational system, L2SCA, and also briefly mentions the characteristics of some other computational systems.

3 Theoretical background – the object

3.1 Definition, form and position

Generally, the object may be defined as a clause element complementing the verb and, in some cases, also the adjective (Dušková et al. 2012: 423)¹². The type of the verb (i.e. copular, intransitive, transitive) determines whether an object is an obligatory or an optional element in the clause (Ibid.). There are two tools that can be used to determine an object in a clause, namely the question test and the passive test. The question test relies on eliciting the object by questions starting with *what* and sometimes also *who*, *whom* (Ibid.: 424). The test includes the fronting of the *wh*- word and subject-operator inversion (Quirk et al. 1985: 728). The passive test is centred around the claim that "[t]he object of an active clause may generally become the subject of the corresponding passive clause" (Quirk et al. 1985: 727), so when applying this test, the correspondence between the active and the passive version of the sentence shows both the subject and the object. Both of these tests are especially helpful when distinguishing between objects and adverbials as the passive test cannot be applied to adverbials, and the questions used to elicit adverbials do not ask *what* but *when*, *where*, *why*, *how* etc. (Dušková et al. 2012: 424). An issue arises when the question "how much" is considered as both objects and adverbials may be elicited by it, and this shows that it is sometimes possible that the postverbal member in a clause possesses some of the qualities of both of the elements (i.e. object, adverbial). There is no simple resolution to this issue apart from the realisation that the transition between the two elements is relatively smooth; thus, there is no fine line to be drawn between them (Ibid.: 425).

As already indicated above, an object typically follows the subject (S) and the verb (V); thus, it may be labelled as a postverbal element in the clause (Quirk et al. 1985: 726). However, in certain cases, the object becomes a preverbal element, namely *wh*-questions that ask about the object and exclamative sentences starting with *what* (Dušková et al. 2012: 424). Regarding the form of the object, pronouns with different case forms require the objective form (Quirk et al. 1985: 727); hence, the only difference occurs in personal pronouns *I*, *he*, *she*, *we*, *they* (*me*, *him*, *her*, *us*, *them*) and in interrogative and relative *who* (*whom*)

¹² „Předmět je rozvíjející větný člen, který rozvíjí sloveso, v některých případech též adjektivum.“ (Dušková et al. (2012: 423), translation Ježková)

(Dušková et al. 2012: 423)¹³. Moreover, Quirk et al. (1985: 726-7) state, "[i]f an object is coreferential with the subject, it usually requires a reflexive pronoun which agrees with the subject in person and, where relevant, in number and gender." and similar rules apply to emphatic genitive (e.g. *my own*); therefore, these cases should also be kept in mind in the question of the form of the object. There are no other formal changes in pronouns or nouns; therefore, their position is the only way to differentiate between the object and the subject (Dušková et al. 2012: 423). What needs to be pointed out is that most of the above-described characteristics apply mainly to the general type of object – the direct object, and there might be slight differences in other object types, as will be shown below.

3.2 Types of objects

The approach adopted in this thesis will consider three types of objects, viz. direct (O_d) indirect (O_i) and prepositional object (O_p) (cf. Biber et al. (1999), Dušková et al. (2012), Quirk et al. (1985)). This is not a generally accepted division. Huddleston and Pullum (2002), for example, do not operate with the term "prepositional object", and use the term "indirect object" to refer both to what is considered an indirect object by Biber et al. (1999) and to the corresponding prepositional phrases. The fact that O_i and O_p are similar in that "they require a mediating element (a preposition or a direct object)" (Biber et al. 1999: 130) may serve as a possible explanation for the discrepancy in the terminology.

3.2.1 Direct object

The direct object (ex. 1a, b) is "a grammatically distinct element of clause structure which in canonical agent-patient clauses expresses the patient role" (Huddleston and Pullum 2002: 245). Nevertheless, the scope of semantic roles that direct objects may take is rather broad. It includes the roles of affected/patient (as mentioned in the definition), resultant, locative, instrumental, measure, cognate, eventive, dummy/empty object (*it*), anticipatory object (*it*), source/stimulus, and agent (Biber et al. 1999: 127-8; Dušková et al. 2012: 425-9). Formally, direct objects occur only with transitive verbs, irrespective of the subtype (i.e. monotransitive, ditransitive, complex transitive) (Biber et al. 1999: 126).

¹³ „Formálně je předmět v angličtině odlišen pouze u zájmen osobních (me, him, her, us, them) a tázacího a vztažného who (whom).“ (Dušková et al. (2012: 423), translation Ježková)

As far as the complementation of an adjective by O_d is concerned, the occurrence of this phenomenon is rare, for example, after *worth* or *like* (Dušková et al. 2012: 444).

- (1) a. *Pat overlooked the error.* (Huddleston and Pullum 2002: 246)
- b. *They amuse me.* (Quirk et al. 1985: 726)

Another point worth mentioning is the terminology connected to directness (i.e. direct and indirect object). The explanation behind this terminology can be seen in ditransitive clauses, as it is believed that "the O_d argument is more directly affected or involved in the process than the O_i argument" (Huddleston and Pullum 2002: 245). Generally, it can be claimed that when there is only one object in the sentence, it is typically O_d ; however, there are some exceptions, which will be described in the section below. Lastly, the omission of O_d is rare, and if it occurs, the originally indirect object will be interpreted as the direct object, and this is when another terminology issue comes into place as there is disagreement among authors as to whether the "new" direct object should be labelled as direct object, indirect object or just simply as "object" (Quirk et al. 1985: 727).

3.2.2 Indirect object

As opposed to the direct object, the indirect object (ex. 2a, b), or dative, is very limited from a formal point of view as it occurs solely in the ditransitive pattern together with the O_d (Biber et al. 1999: 128). Huddleston and Pullum (2002: 245) define O_i as "a distinct element of clause structure characteristically associated with the semantic role of recipient". Apart from the prototypical semantic role of a recipient, O_i may also be the affected/patient in cases where the O_d denotes the event, and also benefactive (Biber et al. 1999: 129). The semantic roles of a recipient (which corresponds to a paraphrase with *to*) and benefactive (which corresponds to a paraphrase with *for*) bring about the topic of possible paraphrasing of indirect objects with prepositional ones as, generally, O_i corresponds to the prepositional phrase realising the O_p (ex. 3a, b) (Quirk et al. 1985: 727; Biber et al. 1999: 129). Another essential aspect of O_i is its position in a clause as the order in ditransitive clauses is fixed – O_d follows O_i ; a switch in the order of these elements means a change in their functions (Biber et al. 1999: 248). Nevertheless, there is one main exception connected to the position of the two objects, and that is when a personal pronoun realises both objects as in this case, their positions are switched, and the O_d precedes O_i (Dušková et al. 2012: 433). Lastly,

if O_i is omitted, this usually does not have any effect on the semantic relations between the other elements (Quirk et al. 1985: 727).

- (2) a. *I gave him* [O_i] *my address* [O_d]. (Quirk et al. 1985: 726)
b. *I wish you* [O_i] *a pleasant journey* [O_d]. (Dušková et al. 2012: 433)
- (3) a. *I'll send Charles* [O_i] *another copy* [O_d]. (Quirk et al. 1985: 727)
b. *I'll send another copy* [O_d] *to Charles* [O_p]. (Ibid.)

A special subtype of the object is the free dative, which is realised by either O_i or O_p with *to* or *for*. The free dative is rather infrequent in English and occurs when the verb does not require a second object (i.e. the verb is not ditransitive in nature). The object, free dative, is added freely to the sentence to express some kind of a participant who either benefits from or is negatively involved in the situation. According to the way the participant is affected, it is possible to distinguish between (a) *dativ commodi* (positively affected participant) and (b) *dativ incommodi* (negatively affected participant). The former is associated with the prepositional phrase with *for*, the latter with the prepositional phrase with *on* (Dušková et al. 2012: 437-9).

3.2.3 Prepositional object

From a formal point of view, the prepositional object occurs with prepositional verbs. It occurs both in the ditransitive and in the monotransitive clause pattern. In the ditransitive pattern, O_d precedes O_p (Dušková et al. 2012: 442).

As opposed to the rareness of complementation of adjectives by the direct object, the prepositional object occurs in this position much more often. Dušková et al. (2012: 444) provide a list of examples of prepositional verbs that tend to complement adjectives: *angry with sb/at sth*, *anxious about sb/sth*, *aware of sth*, *afraid of sb/sth*, *averse to sth*, *bad at sth*, *busy with sth*, *capable of sth*, *certain of (about) sth*, *(in)compatible with sth*, *conscious of sth*, *content with sth*, *dependent on sth/sb*, *devoid of sth*, *due to sb/sth*, *familiar with sth*, *famous for sth*, *(be) fond of sth*, *free from sth*, *good at sth*, *guilty of sth*, *innocent of sth*, *intent on sth*, *keen on sth*, *responsible for sth*, *severe on sb*, *subject to sth/sb*, and *sure of (about) sth*.

3.3 Realisation forms of objects

The realisation forms of objects vary according to the type of object, i.e. direct, indirect and prepositional, which will be described below.

Firstly, the prototypical realisation form of a direct object is an NP. The direct object can also be realised by a subordinate clause, both finite (nominal clause) and nonfinite (infinitive, gerund) (Biber et al. 1999: 126; Dušková et al. 2012: 430). Secondly, the range of realisation forms of an indirect object is much more limited than O_d . Again, the most common realisation form is an NP; however, the only possible type of subordinate clause to realise the object is the *wh*-clause (finite nominal relative clause) (Biber et al. 1999: 128). Lastly, the prototypical realisation form of a prepositional object is a prepositional phrase (PP), which may consist of (a) a preposition + NP; (b) a preposition + a finite clause (*wh*-clause only¹⁴); (c) a preposition + a gerund (i.e. *-ing* clause); or (d) *for* + infinitive (Dušková et al. 2012: 439-41).

Table 2 summarises and exemplifies possible verb complementation types. This thesis does not concern copular verbs and the realisation forms of their complementation.

Variants	Example
Monotransitive (Type SVO)	
[A1] Noun phrase as O (with passive)	<i>Tom caught the ball.</i>
[A2] Noun phrase as O (without passive)	<i>Paul lacks confidence.</i>
[A3] <i>That</i> -clause as O	<i>I think that we have met.</i>
[A4] <i>Wh</i> -clause as O	<i>Can you guess what she said?</i>
[A5] <i>Wh</i> -infinitive as O	<i>I learned how to sail a boat.</i>
[A6] <i>To</i> -infinitive (subjectless) as O	<i>We've decided to move house.</i>
[A7] <i>-Ing</i> clause (subjectless) as O	<i>She enjoys playing squash.</i>
[A8] <i>To</i> -infinitive (with subject) as O	<i>They want us to help.</i>
[A9] <i>-Ing</i> clause (with subject) as O	<i>I hate the children quarrelling.</i>
Complex transitive (Types SVOC and SVOA)	
[B1] Adjectival C_o	<i>The music drives me mad.</i>

¹⁴ There is one exception in which the prepositional object may be realised by a clause starting with *that* and that is in clauses beginning with *in that*, *save that* or *except that* (Dušková et al. 2012: 441).

[B2] Nominal C _o	<i>They named the ship 'Zeus'.</i>
[B3] O + adverbial	<i>I left the key at home.</i>
[BC4] O + <i>to</i> -infinitive	<i>They knew him to be a spy.</i>
[B5] O + bare infinitive	<i>I saw her leave the room.</i>
[B6] O + <i>-ing</i> clause	<i>I heard someone shouting.</i>
[B7] O + <i>-ed</i> clause	<i>I got the watch repaired.</i>
Ditransitive (Type SVOO)	
[C1] Noun phrases as O _i & O _d	<i>They offered her some food..</i>
[C2] With O _p	<i>Please say something to us.</i>
[C3] O _i + <i>that</i> -clause	<i>They told me that I was ill.</i>
[C4] O _i + <i>wh</i> -clause	<i>He asked me what time it was.</i>
[C5] O _i + <i>wh</i> -infinitive clause	<i>Mary showed us what to do.</i>
[C6] O _i + <i>to</i> -infinitive	<i>I advised Mark to see a doctor.</i>

Table 2: Complementation of transitive verbs (adapted from Quirk et al. 1985: 1171)

3.4 Important related notions

Several terms have been or will be used in this thesis; thus, their descriptions and definitions will be given in this section.

3.4.1 Noun phrase (NP)

Biber et al. (1999: 574) define a noun phrase as "a cover term for two major types of construction: noun-headed phrases and pronoun-headed phrases". The structure of an NP differs slightly according to the type of head. If the NP is headed by a noun, the only obligatory parts of the phrase are the determiner and the head noun. In addition, there are two optional components, namely, premodification and postmodification (and complementation). On the other hand, if the NP is pronoun-headed, the only obligatory component is the head pronoun, and the remaining three (same as in the noun-headed phrase) are optional (Ibid.).

3.4.2 Determination, modification, complementation

Determination is a term that "may be used for the function words and (sometimes) phrases which, in general, determine what kind of reference a noun phrase has" (Quirk et al.

1985: 64). Determination is obligatory in NPs headed by nouns, even though it may be expressed by a zero determiner.

As opposed to determination, modification is not primarily obligatory. There are two types of modification: (a) premodification (modifiers precede the head) and (b) postmodification (modifiers follow the head) (Ibid.: 65). The main difference between premodifiers and postmodifiers is that the former are often much shorter and have a tendency to use fewer words; moreover, the two vary in the degree of explicitness, where premodifiers are less explicit in their description of the relationship between the modifier and the head (Biber et al. 1999: 588). Premodification may be realised by four major structural types, namely by (a) a general adjective; (b) an *ed*-participial modifier; (c) an *ing*-participial modifier; and (d) a noun (Ibid.). Moreover, there are three minor structural types of premodification: (a) genitive; (b) adverb and other phrases; and (c) sentence (Quirk et al. 1985: 1322). There are several structural types of postmodification; the four major ones are (a) finite clauses (relative clauses); (b) non-finite clauses (*to*-clauses, *ing*-clauses, *ed*-clauses); (c) noun complement clauses; and (d) prepositional phrases. Moreover, a head may also be postmodified by (a) an adverb (phrase); (b) an adjective (phrase); and (c) an emphatic reflexive pronoun in apposition (Biber et al. 1999: 604).

The last notion to be mentioned is complementation, defined by Quirk et al. (1985: 65) as "the function of a part of a phrase or clause which follows a word, and completes the specification of a meaning relationship which that word implies". Regarding the obligatoriness of complementation, it may be either obligatory or optional (Ibid.).

3.4.3 Apposition

Apposition is a coordinating relationship between two clauses or phrases characterised by the fact that the relationship members have the same syntactic function. Considering apposition to have characteristics of coordination may be supported by the fact that when one member of the relation is removed, the syntactic function of the remaining member remains the same (Dušková et al. 2012: 498). Moreover, Quirk et al. (1985: 1301-2) add: "Apposition resembles coordination in that not only do coordinate constructions also involve the linking of units of the same rank, but the central coordinators *and* and *or* may themselves occasionally be used as explicit markers of apposition." The relationship between the

members of an apposition may be either restrictive or non-restrictive (Biber et al. 1999: 638). Though it may sometimes be challenging to recognise the appositive relationship, there are some explicit indicators of apposition that may help with the identification: *and, as, as follows, chiefly, e.g., especially, for example, for instance, i.e., in other words, in particular, in simple(r) words/terms, in technical terms, included, including, like, mainly, mostly, namely, notably, or, or better, or rather particularly say, (more) simply, such as, technically, that is, that is to say, to wit, viz* (Dušková et al. 2012: 499, Quirk et al. 1985: 1307).

4 Material and method

4.1 Material

This thesis draws on three separate corpora: the Czech subcorpus of The Varieties of English for Specific Purposes dAtabase (VESPA_CZ) used as an L2-novice source; the British Academic Written English Corpus (BAWE) used as an L1-novice source; and the L1-expert corpus (created through Sketch Engine).

4.1.1 The Varieties of English for Specific Purposes dAtabase (VESPA_CZ)

The Czech subcorpus of The Varieties of English for Specific Purposes dAtabase (VESPA_CZ) is a learner corpus compiled of academic texts written by Czech university students. The VESPA_CZ subcorpus is a part of the project VESPA, performed by the Centre for English Corpus Linguistics at the University of Louvain in Belgium. This international project aims “to build a large collection of disciplinary writing by L2 English university students across registers and disciplines”¹⁵. The subcorpus comprises texts collected in 2019-2022 at the Faculty of Arts, Charles University, Prague (BA and MA programmes ‘English and American Studies’, ‘Anglophone Literatures’ and ‘English Language’) and the Faculty of Business Administration, University of Economics and Business, Prague (BA programme ‘Arts Management’)¹⁶. The VESPA_CZ subcorpus was accessed via KonText, and for the purposes of this thesis, a subcorpus was created and restricted to literature courses. The created subcorpus comprises 80 texts, 145723 tokens, 127556 words and 4901 sentences.

4.1.2 British Academic Written English Corpus (BAWE)

The British Academic Written English Corpus (BAWE) is a learner corpus of proficient university-level student writing. The creation of this corpus resulted from the research project *An Investigation of Genres of Assessed Writing in British Higher Education*, which was carried out in 2004-2007 at Oxford Brookes, Reading and Warwick universities (Heuboeck et al. 2010: 4). The texts in BAWE come from 35 distinct disciplines (e.g. Archaeology, English, Health, Psychology, Mathematics, Business, Sociology) which

¹⁵ (<https://uclouvain.be/en/research-institutes/ilc/cecl/vespa.html> Accessed 20 June 2024)

¹⁶ (https://wiki.korpus.cz/doku.php/en:cnk:vespa_cz Accessed 20 June 2024)

are divided into four major disciplinary groups, namely, Arts and Humanities, Life Sciences, Physical Sciences and Social Sciences (Ibid.: 5). The corpus is made up of 2897 texts and over 6.5 million words in total written by 1039 students, and it features several genres (e.g. essay, literature review, proposal) (Ibid.). For the purposes of this thesis, a subcorpus was created by restricting the BAWE to the discipline of English within the Arts and Humanities (AH) disciplinary group. The created subcorpus comprises 86 texts, 256354 tokens, 213771 words and 9073 sentences.

4.1.3 L1-expert Corpus

The corpus for the L1-expert texts was created using academic texts from three different journals: English Literary Renaissance, Renaissance Studies, and Shakespeare Quarterly. English Literary Renaissance journal publishes articles on Tudor and Stuart literature (e.g. Spenser, Shakespeare, Donne)¹⁷. Renaissance Studies (RS) is a journal that focuses on describing the Renaissance period and literature from several perspectives, including historical, cultural, and religious backgrounds¹⁸. Shakespeare Quarterly (SQ) publishes essays, notes and book reviews revolving around Shakespeare and his work. The articles in SQ try to read Shakespeare from the points of view of queer and gender studies, premodern critical race studies, textual studies or theatre history¹⁹. These three specific journals were selected because their topics are similar to those in the other two corpora. The corpus was created via Sketch Engine and comprises 34 texts, 307009 tokens, 256793 words, and 8581 sentences.

4.1.4 Register characteristics of written academic texts

Biber and Conrad (2019: 31) define register as “a language variety associated with a particular situation of use” and mention the three main parts of register analysis: (a) the situational characteristics; (b) linguistic features; and (c) communicative functions. To conduct such an analysis, Biber and Conrad (2019: 40) introduce a framework of situational parameters of different registers and genres. Using this framework, it is possible to determine

¹⁷ (<https://www.journals.uchicago.edu/journals/elr/about> Accessed 20 June 2024)

¹⁸ (<https://www.jstor.org/journal/renastudies> Accessed 20 June 2024)

¹⁹ (<https://www.folger.edu/research/shakespeare-quarterly/> Accessed 20 June 2024)

the situational characteristics of written academic texts (cf. Table 3) in the three analysed subcorpora.

Participants	Addressor(s)	single, academic novice/expert
	Addressee	unenumerated, other
Relations among participants	Interactiveness	no direct interaction
	Shared knowledge	specialist
Channel	Mode	writing
	Specific medium	permanent, printed
Processing circumstances	Production	planned, revised, edited
Communicative purposes	General purposes	inform, explain, describe
	Specific purposes	re-interpret, summarise
	Purported factuality	factual
	Expression of stance	no overt stance
Topic	General topic domain	academic
	Specific topic	literature

Table 3: The situational parameters of the three analysed subcorpora (cf. Biber and Conrad 2019)

In addition, it is possible to describe written academic texts from the point of view of complexity. These texts rely on phrasal embedding as a source of complexity, in comparison to spoken registers, where the reliance is on clausal subordination (Biber and Gray 2010: 18). The purpose of the tendency of academic texts being rather compressed is language economy and easier processing (Ibid.: 11). Another idiosyncrasy of written academic texts is the employment of prepositional phrases as noun post-modifiers. Lastly, it needs to be pointed out that in terms of grammatical characteristics, written academic texts are “dramatically different from all spoken registers and most written registers” (Ibid.: 18).

4.1.5 Learner corpora

The term learner corpora may be defined as “corpora representing written and/or spoken ‘interlanguage’, that is, language produced by learners of that language” (Gilquin 2020: 283). Moreover, Callies (2015: 35) states that “learner corpora provide systematic collections of authentic, continuous and contextualised language use by foreign/second language (L2) learners stored in electronic format”, by which he points out the idiosyncratic nature of learner corpora as opposed to other types of data used in second language

acquisition (SLA) research. The most important merit of learner corpora is that “[t]hey enable the systemic and (semi)-automatic extraction, visualisation and analysis of large amounts of learner data” (Callies 2015: 35).

The establishment of learner corpora took quite a long time, and the collection of the first data to create such corpora began around the 1980s. These tendencies resulted in the establishment of the *International Corpus of Learner English* (ICLE) or the *Longman Learners’ Corpus* (Gilquin 2020: 284). Unlike other regular corpora-related data collections and methodologies, learner corpora have come with many distinctive challenges. The greatest challenge was “to design methods that could help researchers uncover the distinctive characteristics of learner language” (Granger 2015: 7), which forced the linguists to make revisions and readjustments of some of the principal theories and concepts of corpus data collection and analysis (Gilquin 2020: 284). The results of these new tendencies include the creation of new types of corpora (e.g. longitudinal corpora) or new approaches to the collection of metadata, which are often very large in amount. The metadata address either the text itself or the learners (including their mother tongue (L1), type and time of exposure to the target language (L2) or years of study) (Ibid.: 286).

Learner corpora have several specific features, which will be addressed below. Firstly, learner corpora are characteristic in their size, as a majority of them include under one million words, some consisting only of ten thousand words (Gilquin 2020: 284). Secondly, the texts that create learner corpora should be authentic; however, the understanding of authenticity is rather complex and needs to be specifically accustomed, as the texts produced by learners are often directed to fulfil some pedagogical aim; thus, the aim is not solely to reach a communicative purpose (Ibid.: 285). Thirdly, the researchers ought to consider the possibility that “some learner corpora will be more ‘peripheral’” (Ibid.). And lastly, there is a genre-related issue in learner corpora. The range of genres in learner corpora is rather limited because the texts are produced in educational settings; hence, the most common genres that appear in learner corpora are argumentative essays, academic writing, narratives, and interviews (Ibid.).

4.1.6 Academic and linguistic challenges

When focusing on differences between the three analysed groups, it is relevant to focus on the effects of academic and linguistic challenges. The former one, academic challenge, applies to both L1 novice and L2 novice writers and may be defined as “the lack of experience on the part of novice academic writers” (Malá and Raušová 2022: 10). Such a challenge is connected with English for Academic Purposes (EAP) and also with the problems novices deal with when they wish to “locate their writing within the academic community they are entering as inexperienced writers” (Ibid.: 12). The latter one, linguistic challenge, applies only to the L2-novice group as it is caused by them “not being native English writers” (Ibid.: 10) and their “need to gain control of the English language at a high level of proficiency” (Ibid.: 11), and also by having English as a Foreign Language (EFL).

4.2 Method

The analytical part of this thesis combines quantitative (percentages, figures, etc.) and qualitative analysis (interpretation) of the data, relying on corpus-assisted research. The underlying approach of our analysis is Hunston’s (2015, in Biber & Reppen) functional lexico-grammatical view of language, connecting lexical and grammatical features. The method used in the analytical part is based on Granger’s (2015) Contrastive Interlanguage Analysis (CIA), focusing on a twofold comparison of a language: (a) with native language (novice vs. expert) and (b) with different learner language (L1 vs. L2) (Granger 2015: 8). In addition, this thesis adopts Biber et al.’s (2011, 2023) distinction of clause elements according to their structural type (finite dependent clauses, non-finite dependent clauses, dependent phrases) and syntactic functions (clause constituent, NP constituent, other-phrase constituent).

4.2.1 Procedure

The procedure in the analysis was similar in the three studied corpora (L1-expert corpus, BAWE, VESPA_CZ). The steps that have been taken are described below.

Generally, the query which was used as the initial step in order to select individual random sentences was the same in all three corpora: “<s> [word=".+"]”, with ‘word’ as the ‘default attribute’. In both corpora analysed via Sketch Engine (i.e. L1-expert and BAWE), the

concordance was ‘advanced’, and the ‘query type’ was set to CQL. Regarding the subcorpus accessed via kontext, ‘advanced query’ was selected.

There were also specific restrictions resulting from the creation of the analysed subcorpora. In the L1-expert corpus, no restrictions needed to be selected as the whole corpus was created by the authors of this thesis. Within BAWE, the query was restricted to the ‘AH’ discipline group (area) and the ‘English’ discipline. In VESPA_CZ, the search was restricted to the ‘literature’ courses.

The query was run in each of the subcorpora, and the matches were shuffled. Then, a thorough analysis of the matches was held, selecting 50 example sentences with objects in the main clause from each of the subcorpora. Examples involving objects in quotation, passive sentences, objects only in subordinate clauses or emphatic structures²⁰ (e.g. cleft and pseudo-cleft sentences) were not included in the analysis. The analysis then consisted in classifying 50 random examples from each corpus according to the criteria outlined in the theoretical part of this thesis. The examined criteria were: main clause verb (predicate); clause elements (clause pattern); transitivity of the predicate verb; type of object (O_d, O_i, O_p); length of the object phrase or clause, structural type; and levels of embedding. Within the analysis of main clause verbs, the verbs were assigned their ranks according to *new-GSL*²¹. Lastly, the 50 examples from each corpus were analysed in relation to Biber et al.’s (2011) hypothesised developmental stages.

²⁰ From formal point of view, such sentences involve a main *it*-clause and a subordinate *that*-clause with an object, thus emphatic structures cannot be considered as having objects in the main clause.

²¹ „The new-GSL (New General Service List), developed in 2013 by Vaclav Brezina and Dana Gablasova of Lancaster University, is a list of 2494 words which occur frequently in a range of texts.“ (<https://www.capfoundation.com/vocab/general/newgsl/?type=freq#newgsllist> Accessed 24 June 2024)

5 Analytical part

5.1 Main clause verbs

The main clause verbs in each of the three studied text samples (i.e. L2-novice, L1-novice and L1-expert) can be described according to several parameters, some of which will be elaborated on below.

Firstly, it is possible to look at the different types of verbs²². There were 41 types of verbs in the examined L2-novice texts. Some of the verbs occurred more than once, namely *say* (3 instances), *know* (2 instances) and *have* (3 instances). The total number of types of verbs in L1-novice texts was similar to that of L2-novice texts; more precisely, there were 40 types of verbs in the L1-novice corpus data. However, there were more multiple occurrences of the verbs than in L2-novice texts. All of the following words occurred twice in the studied sample of L2-novice texts: *believe*, *lead*, *make*, *mean*, *see*, and *use*. In the L1-expert texts, the authors used 44 different verbs. In this case, three verbs occurred twice: *note*, *play* and *say*. Overall, there is no substantial difference in the number of verbs the authors used in the three studied groups. However, it is visible that L1-expert writers use a slightly wider variety of verbs, and surprisingly, L1-novice writers tend to use the same verbs more than L2-novice writers. Despite the fact that our sample is not so large, it can be asserted that novice writers generally opt for a narrower range of verbs than expert writers; moreover, it is also clear that the verbs used by novice writers are repeated more frequently as opposed to those used by expert writers.

Secondly, the verbs can be described from the point of view of their lexical sophistication. Such a description was done using the New General Service List (*new-GSL*). The lower the rank in the *new-GSL*, the more frequent and ordinary the examined word (in our case, verb) is. Some of the verbs that appeared in our samples are not a part of the *new-GSL* list; thus, these verbs were considered infrequent, and they were all given the rank of 2500 (the highest rank in the *new-GSL* is 2491). All of the verbs in each group were assigned a ranking

²² When analysing the verbs, it is essential to verify that the verb in question does not appear more than once in the analysis only because it is complemented by two objects (i.e. it is a ditransitive verb). If such a verb was present in the analysis, the multiple occurrences were counted as one type of the verb. On the contrary, if a verb occurred as a predicate in more than one matrix clause, all of its occurrences were counted and regarded as one instance.

according to the new GSL, and their means and medians were then calculated. The verb rank can be seen in Figure 1. The mean of L2-novice writers is 834, so it can be said that they tend to use more frequent general verbs. The median of L2-novice writers is 619, indicating that the used verbs have a similar and higher frequency (lower ranking in the *new*-GSL). Moreover, the range of ranks of L2-novice texts is the narrowest, indicating that the writers in this group are likely to use a lower variety of verbs, which are often not so advanced. As for L1-novice writers, the mean of the ranks is 952, slightly higher than that of L2-novice writers; hence, the trend here is to use more sophisticated verbs. As expected, the texts of L1-expert writers display the highest level of sophistication of verbs, with a mean of 1199. Furthermore, the median also has the highest figures (1037), which shows that the range of verb usage by L1-expert authors is the most extensive. In general, it can be spotted that the mean of the ranks gradually increases from L2-novice being the lowest to L1-expert being the highest. Surprisingly, the difference between the verb ranks of L2 and L1 novice writers is relatively considerable, where the range of verb usage of L1-novice writers is much broader than that of L2-novice writers.

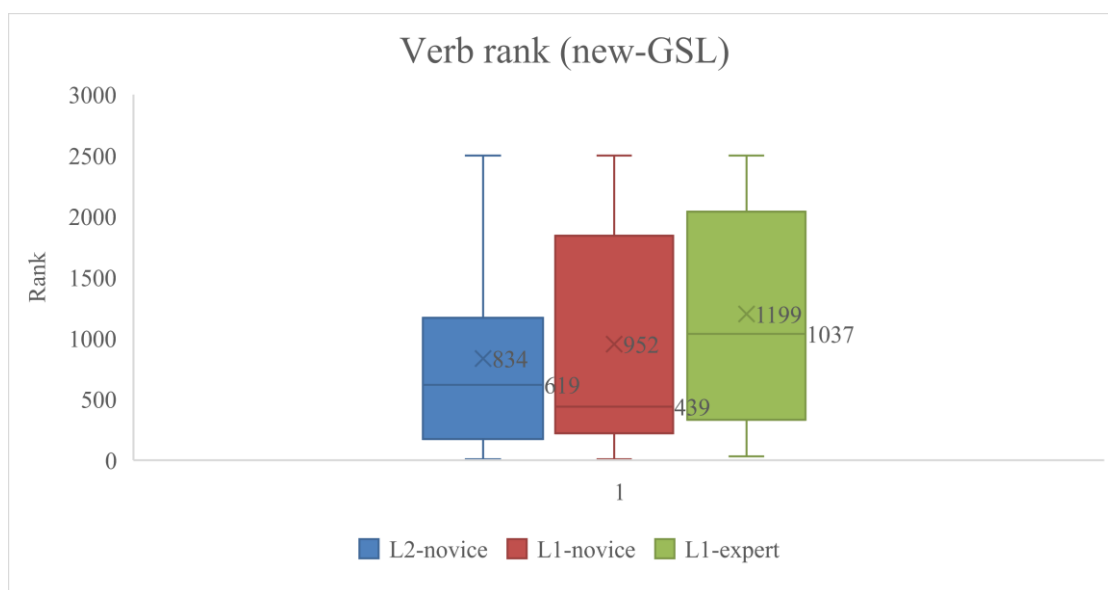


Figure 1: The verb rank according to *new*-GSL in all three subcorpora.

The last area directly connected to the analysis of main clause verbs is the question of overlap. There were eight overlapping verbs between L2-novice and L1-novice texts,

namely, *describe* (398²³), *have* (8), *lead* (214), *make* (35), *see* (57), *suggest* (297), *support* (397) and *use* (76). It is obvious that all of these verbs are considerably frequent as their ranks are low. This may imply that novice writers share the favour in using more general and rather rudimentary verbs. The overlap between verbs in L2-novice and L1-expert samples is composed of the following eight verbs: *make* (35), *paint* (2252), *play* (230), *present* (559), *provide* (177), *say* (32), *see* (57) and *urge* (1854). It can be stated that the verbs that L2-novice and L1-expert samples share belong to the academic register, and this may serve as an explanation for the higher ranks of some of the verbs (e.g. *present*, *urge*). Interestingly, one of the overlapping verbs in the two groups was the verb *paint*, which has a very high rank. The occurrence of this verb in the texts is probably due to the topics they revolve around (i.e. Renaissance literature). Regarding the area of verb overlap in L1-novice and L1-expert texts, this area is the smallest of the three. There were only six overlapping verbs in these texts, and these were *attempt* (1182), *highlight* (1698), *include* (176), *make* (35), *see* (57) and *seek* (522). Under closer examination, it can be seen that the overlapping verbs *attempt*, *highlight* and *seek* are verbs of academic language, and they are rather infrequent. Moreover, the list of overlapping verbs includes the verb *include*, which is profoundly important in expert writing. Overall, the very general and frequent verbs *make* (35) and *see* (57) were the only two verbs that occurred in all three subcorpora, which may indicate the tendency of all writers to opt for the use of simpler vocabulary if appropriate. In addition to the monotransitive pattern (SVO), both groups of novice writers used the verb *make* in a complex-transitive construction (SVOC_o).

5.2 Transitivity, verb valency

As the main focus of this thesis is the object and its realisations in three different types of texts, the transitivity of the main clause verbs was analysed. Table 4 summarises the distribution of transitivity in all three analysed subcorpora. It can be seen that over 82 % of verbs (113 instances) were monotransitive, around 8 % (11 instances) were ditransitive, and the remaining 10 % (14 instances) were complex transitive. Out of the 113 monotransitive

²³ The numbers in brackets represent the rank according to *new-GSL*.

verbs, only 11 were complemented by a prepositional object; the remaining 102 were complemented by a direct object.

type of verb	L1-expert	L1-novice	L2-novice	total
monotransitive	41	37	35	113
ditransitive	3	4	4	11
complex transitive	3	5	6	14
total	47	46	45	138

Table 4: Types of verbs according to their transitivity in the chosen samples (the total numbers of verbs are lower than 50 in each sub-corpus since the ditransitive verbs were counted only once in this table, even though they are complemented by two objects)

In the case of ditransitive verbs, there were two types of patterns: (a) with an indirect and a direct object (SVO_iO_d); and (b) with a direct and a prepositional object (SVOO_p) (cf. Table 5). The numbers of ditransitive patterns are too low to draw any conclusions beyond the fact that the pattern with a prepositional object is more frequent than the SVO_iO_d pattern in all sub-corpora. In the works of both groups of L1 writers, there were attested instances of reversed order of the direct and prepositional object due to the ‘end-weight’ principle, i.e. the tendency for the more complex, heavier, clause elements (O_d) to move towards the end of the clause (ex. 1²⁴).

- (1) [...] this does not impress **upon him** (Op) **a profound respect for authority in general, or Angelo in particular** (O_d) (L1-expert)

pattern	L1-expert	L1-novice	L2-novice	total
SVO _i O _d	1	 	1	2
SVOO _p	2	4	3	9
total	3	4	4	11

Table 5: Types of clause patterns in ditransitive verbs in the samples²⁵

Regarding complex transitive verbs, it is evident (cf. Table 4) that their usage is higher in both novice groups. More accurately, the pattern SVOC_o is very common. All five cases of complex transitive verbs in L1-novice writers (ex. 2) and five of the seven examples in the L2-novice group (ex. 3) feature the pattern with an object complement.

²⁴ In the analytical part of the thesis the numbering of the examples starts from 1 again.

²⁵ The crossed-out cells in the tables stand for zero occurrences of the feature at hand.

- (2) Accepting Humm’s belief that; ‘gender is constructed through language and is visible in writing style;’ using direct, pithy end-stopped sentences rarely coloured by emotion make **Movern** appear more masculine, as conventional male language is seen as ordered and rational. (L1-novice)
- (3) In other words , she regards **marriage** just one of the few occupations a woman is offered during Victorian times. (L2-novice)

5.3 Length of objects

One of the factors that may suggest an increase in complexity is the increase in the length of a clause element. Apart from many other things, the length of the structures realising objects differs among the three analysed groups (cf. Figure 2).

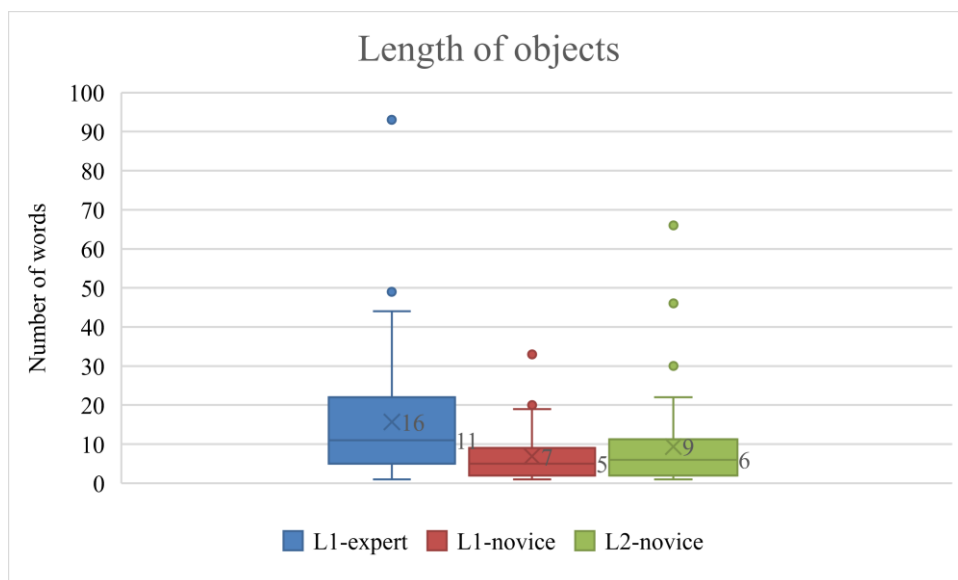


Figure 2: The number of words in the object in all three subcorpora

The average length of objects in L1-expert is around 16 words, in L1-novice around 7 words, and in L2-novice around 9 words. Moreover, when the medians of all three groups were counted, the results were as follows: 11 words for L1-expert, 5 words for L1-novice, and 6 words for L2-novice. The figures of medians show that expert writers tend to use longer structures for objects, as half of the studied samples were composed of more than eleven words, whereas the figures of both novice groups are around half the size. Moreover, novice writing displays a narrower range of values, i.e. there is not much

variation among individual novice writers with respect to the length of the object. The minimum (1 word) is the same for all the groups, while the maxima and outliers differ significantly. The longest object used by L1 experts was realised by a finite clause composed of 93 words. In L1-novice texts, the longest object comprised “only” 33 words. Both of the longest structures in the L1 texts involved only two levels of embedding. The longest structure in the L2-novice group consisted of 66 words with five levels of embedding.

Another interesting length-related feature is the most frequently used length of structures. As expected, the shorter the structure, the more likely it is to be used more frequently; hence, structures composed of only one word (i.e. pronouns, proper nouns, abstract uncountable nouns or nouns in plural) were the most commonly used type among all three groups. A surprising finding comes with the usage of 2-word structures (i.e. usually simple noun phrases consisting of a determiner and a head noun, or prepositional phrases with a preposition complemented by a simple noun phrase comprising merely the head noun) as expert writers used such structure only once, whereas it was used nine times by L1-novice writers and seven times by L2-novice writers. This finding goes along with the tendency of L1 experts to use longer structures in general. Furthermore, when 3-word structures were analysed, it was discovered that neither of the L1 groups uses them significantly, while L2 novices used such structures six times. A possible explanation for this is that L2-novice writers tend to use premodification by adjectives as the only source of complexity in their texts.

5.4 Types of objects

Generally, direct objects (84 %) predominated in the studied sample, followed by prepositional objects (15 %), and the occurrence of indirect objects was very sparse (1 %) (cf. Table 6).

object	L1-expert	L1-novice	L2-novice	total
direct	43	41	42	126
indirect	1	 	1	2
prepositional	6	9	7	22
total	50	50	50	150

Table 6: Types of objects in the sample

5.4.1 Direct object

As already shown in Table 6 above, direct objects present a vast majority in the studied sample, and unlike the other types of objects, their distribution among the analysed subcorpora is relatively even. In terms of transitivity, direct objects occurred with monotransitive verbs (100 instances), ditransitive verbs (11 instances) and complex transitive verbs (15 instances).

structural type	L1-expert	L1-novice	L2-novice	total
noun phrase (NP)	27	28	29	84
finite clause	11	8	9	28
non-finite clause	5	5	2	12
finite clause + NP			1	1
NP + finite clause			1	1
total	43	41	42	126

Table 7: Structural types of direct object in the sample

As opposed to the other types of objects, direct objects are realised by several structural types (Table 7), most typically by noun phrases (84 instances), followed by finite clauses (28 instances) and non-finite clauses (12 instances). There were two special cases of realisation forms of direct objects, both featuring the coordination of a noun phrase with a finite clause, each in a different order (ex. 4), and surprisingly, both appeared in L2-novice texts. Regarding the differences connected to structural types in each of the analysed groups, the usage of noun phrases as a realisation form for direct objects was similar in all three groups (27 instances in L1-expert, 28 instances in L1-novice and 29 instances in L2-novice). It is remarkable that L1-expert writers used the inherently complex structural type of finite clauses the most (11 instances), which confirms the existence of the academic challenge. The most notable difference between the groups in the structural types of direct objects is represented by non-finite clauses, as both L1 groups used them 5 times, whereas L2-novice writers used non-finite clauses only twice. This may be a result of the linguistic challenge L2-novice writers are faced with, and perhaps the impact of their L1 Czech, where non-finite clauses are used less frequently than in English.

- (4) The fifth stanza describes mostly **the human nature** and **that most people would prefer receiving love and not have to give anything in return.** (L2-novice)

The following paragraphs deal with the various realisations of the noun phrases as structural types of direct objects (Table 8). It is clear that complex noun phrases account for the largest part of noun phrase realisations. In the papers written by L1-experts, they were used 3.8 times more frequently than simple noun phrases. The ratio between simple and complex noun phrases is more balanced in novice writing. What is surprising, however, is that L2-novice writers use complex NPs more often than L1-novice writers. The fact that both novice groups use simple NPs much more frequently than experts confirms that novice writers are still learning the strategies in academic language; thus, they are subjected to the academic challenge. Apart from using complex noun phrases to a larger extent, L1-experts differ from novice writers in employing multiple coordinated objects more frequently (the numbers, however, are too low to allow any conclusions).

NP structural type	L1-expert	L1-novice	L2-novice	total
NP complex	19	11	16	46
NP simple	5	16	11	32
NP complex COORD NP complex	1	1	1	3
NP simple COORD NP complex	1			1
NP proper COORD NP proper			1	1
NP simple COORD quote	1			1
total	27	28	29	84

Table 8: Realisations of noun phrases in the sample

The sources of complexity of complex noun phrases reside in (a) premodification by an adjective (ex. 5a) or an NP (ex. 5b); (b) postmodification by an adverb (ex. 6a), non-finite and finite clauses (ex. 6b, c) or a PP (ex. 6d); and (c) combination of premodification and postmodification (by finite clauses or PPs) (ex. 7a-d).

L2 novice writers frequently use noun phrases whose complexity consists merely in the presence of premodifiers (type a, 37 % of complex NPs) or merely postmodifiers (type b, 32 % of complex NPs). Premodification is realised by adjectives, and postmodification is typically achieved by *of*-prepositional phrases. One of the specifics of this group of writers is the usage of premodification only by an adjective as a source of complexity since it does not appear in L1-novice texts at all, and L1-expert writers used it only once.

L1 novice writers differ from L2 novice writers in relying most frequently on complex noun phrases with postmodifiers (type b, 42 % of complex NPs) or a combination of premodification and postmodification (type c, 58 % of complex NPs.)

Similarly to L1 novice writers, L1 experts use complex noun phrases of the types (b) and (c) most frequently. The complexity of these phrases, however, appears to be of a different type than in the essays written by L1 or L2 novice writers. Complex NPs in L1 expert writing comprise postmodification with heavy embedding.

- (5) a. His name fits **his arrogant personality**. (L2-novice)
b. Resistance theory only occasionally penetrated **the 'good wife' ideology** [...]. (L1-expert)
- (6) a. For the involvement of Volpone's role, he is not able to fully enjoy the results of his effort, as there are plenty of interruptions forcing him to always be ready to play his role, thus he gets almost **no time off** to be himself [...]. (L2-novice)
b. He supports this also by referring to the candle carried by Othello which also suggests the "church' tableau' and sees **the imagery used by Shakespeare** [...]. (L1-novice)
c. Hereby he supports [**the notion [that women are more deceitful and merely beings[driven by their primal urges] [if not controlled by an outside party]] and [need to be treated like property.]**] (L2-novice)
d. Despite these concerns, however, alehouse regulation was scant and attempts at enforcement left **many records of overt resistance** to the law. (L1-expert)
- (7) a. Her demonic capabilities are thoroughly compelling and exemplify a **woman's capacity for ingenuity**. (L1-novice)
b. The three lines **perform a capsule survey of the English countryside, from its heights to its delimiting shores**. (L1-expert)
c. Their performances display **their mutual affection which culminates in the dance**. (L1-novice)
d. The finale of the story contains **the absolute extreme volume of emotions** [...]. (L2-novice)

What should also be mentioned is that the above-mentioned examples and descriptions deal solely with the first and second levels of embedding in the complex NP, while some of the direct objects in the texts display many more levels of embedding, contributing to the complexity of the whole structure. Two examples of this kind will be presented, one from L2-novice texts and one from L1-expert texts. In example (8), the whole realisation form of the object is composed of 66 words, with NP being on the 1st level, postmodification by a PP on the 2nd level, postmodification by two coordinated finite clauses on the 3rd level, further development of the phrase by an adverbial non-finite clause on the 4th level and postmodification by two coordinated finite clauses on the 5th and last level. Now, considering example (9), the structure is composed of 25 words; the first two levels are the same as in example (8), the 3rd level of embedding includes postmodification by a finite clause, the 4th level involves postmodification by a non-finite clause and the 5th level is realised by postmodification by a non-finite relative clause.

- (8) The clear and concise structure matches [**the argument [of the poem], [[which is introduced in the first stanza] and [elaborated upon in the following stanzas]], [listing the manners [[in which the lyrical subject has tried to vow the lady - “serving still”, “my goodwill” (stanza 1, lines 1 and 3), “all my smart” (stanza 3, line 2) as well as his “pain” (stanza 2, line 2)] and [which have all been ignored.]]]**] (L2-novice)
- (9) Cynthia Herrup also notes [**differences [in the rate [in which execution was ordered] [depending upon both the type and value of goods [stolen among charges more serious than theft.]]]**] (L1-expert)

When describing the levels of embedding, a majority of postmodification within the complex noun phrase is phrasal, thus confirming Biber et al.'s (2011) claims that written language tends to be more phrasal than clausal. The confirmation in our case is even deeper as it shows that the sources of complexity are phrasal even on those levels that are not completely at the surface (mainly the first three levels), i.e. on the embedded levels. However, when the complex phrase is examined very thoroughly, the deeper levels are always occupied by clausal structures. Thus, it can be presumed that the deeper the level of embedding, the bigger the probability of clausal features to occur.

The second most commonly employed structural type of direct object was finite clauses, and these will be described in detail in the following paragraphs.

finite clause	L1-expert	L1-novice	L2-novice	total
<i>that</i> -clause	7	4	4	15
<i>wh</i> -clause	1	3	2	5
<i>that</i> -clause COORD <i>that</i> -clause	1	1	1	2
<i>wh</i> -clause COORD <i>wh</i> -clause	1	1	1	2
quote (<i>that</i> -clause)	2	1	1	4
total	11	8	9	28

Table 9: Types of finite clauses in the sample

The fact that nominal content *that*-clauses represent a majority of finite clauses in the studied sample is clear from Table 9. An interesting observation is that L1-novice writers did not use either of the coordinated structures, whereas both L1-experts and L2-novice writers used each of the structures once. In connection to *wh*-clauses, there were two examples of nominal content *wh*-clauses, both in L2-novice texts; the remaining *wh*-clauses were nominal relative. In the case of finite clauses, the structures were often not embedded at all, and if they were, the source of complexity was postmodification by a finite or a non-finite clause (ex. 10a, b).

- (10) a. Both formats were implemented expertly by Surrey , but we can see
[that the second version of a sonnet , [that he built] serves his purposes and style of progression through the poem more elegantly.] (L2-novice)
- b. In response to this, I will examine **[how the themes of gender relations, education and use of landscape are used as vehicles [to write tragedy] in George Eliot's *The Mill on the Floss* (1860) and Charles Dickens's *Great Expectations* (1861)]**, focusing specifically on the final sequences of each novel.
 (L1-novice)

The last structural type that is yet to be described is a non-finite clause. Eleven out of twelve non-finite clauses are infinitival, one has a gerund as its predicate. As mentioned at the beginning of the section dealing with direct objects, there were only 2 instances of non-finite clauses realising objects in L2-novice texts. Expert and novice L1 writers used objects realised by non-finite clauses more frequently, 5 times each. The complexity of non-finite

clauses lies in using either a single non-finite clause (8 instances) or two coordinated non-finite clauses (4 instances). It is interesting that using two coordinated non-finite clauses is quite common and occurs in all three analysed subcorpora. When a single non-finite clause is employed, there are usually more embedded structures within it (ex. 11a, b), whereas when the authors use two coordinated non-finite clauses, typically, no other embedding occurs (ex. 12).

- (11) a. Bacchus continues [**to prepare the audience for the tragic drama** [**they should expect** [**to unfold**]]] [...]. (L1-novice)
 b. Yet in his visceral aversion to Lady Politic's unsolicited presence, Volpone fails [**to register** [**that he has in fact been operating according to** [**the central tenet of her advice**]: [**mental well-being demands the pursuit of novelty.**]]] (L1-expert)
- (12) It is introduced to us already in the title of the story - the readers then know **what to expect and look for** [...]. (L2-novice)

5.4.2 Indirect object

Regarding indirect objects, no particular tendencies can be observed due to their low frequency of occurrence, except that they were both realised by a simple noun phrase (ex. 13a, b).

- (13) a. Later on , he finally brings **Volpone** what is supposed to be considered his most precious possession : Celia . (L2-novice)

5.4.3 Prepositional object

As already implied, prepositional objects were not very frequent in the studied sample. They appear in both monotransitive (13 instances) and ditransitive patterns (9 instances). As far as the length of the structure is concerned, objects complementing monotransitive verbs tend to be slightly longer (4-22 words) than those complementing ditransitive verbs (2-18 words). Regarding main clause verbs and prepositional objects, there are no specific patterns, as none of the verbs is used more than once. The structural type of all of the prepositional objects is a prepositional phrase (PP); however, the complement of the preposition in the PP differs among the studied samples. The most common complementation of a preposition is by a

simple noun phrase or a complex noun phrase. There was one example of complementation by coordinated noun phrases (ex. 14) in L1-expert texts.

- (14) Moreover , short hand reporting accounts **for the differences in speech prefixes and names, widely diverging lineations of verse passages, instances of aural mistakes, and variant stage directions.** (L1-expert)

The complexity of complex noun phrases is based on (a) premodification by one or two adjectives or a genitive (ex. 15); (b) postmodification by a finite or a nonfinite clause or a prepositional phrase (ex. 16a, b); or (c) combination of premodification and postmodification (ex. 17). It needs to be mentioned that when (b) is applied, all of the elements can be further embedded and expanded on, and the occurrence of (c) is present only in L1 texts, both novice and expert.

- (15) Thus the Duke focuses **on Shylock's "strange apparent cruelty"** [...].
(L1-expert)
- (16) a. Poetry provided W . B . Yeats and Anna Akhmatova [**with [a platform where they were able [to employ their extraordinary talent - to create art with an impact.]]]] (; L2-novice)**
- b. Eliot likens their whispering **to " wind in dry grass , " quietly rattling , yet of no real consequence.** (L2-novice)
- (17) In a society where motherhood was synonymous with being a woman, Eulalie completely divorces herself **from society's expectations of womanhood,** [...].
(L1-novice)

5.5 Discussion in relation to developmental stages

This chapter relates the findings of the analysis to Biber et al.'s (2011) hypothesised developmental stages for complexity features (cf. 2.3.3, mainly Table 1). The stages incorporate three main dimensions of complexity applicable to objects (see Table 10), namely (a) clausal realisations of objects (further subdivided into finite and non-finite clauses); (b) phrasal realisations of objects; and (c) appositive and coordinative realisations of objects.

Stage	Grammatical structure		
	clausal realisation of object	phrasal realisation of object	apposition and coordination
1	Finite complement clauses controlled by common verbs (e.g. <i>think, know, say</i>)		
2	Finite complement clauses controlled by a wider set of verbs	Simple phrasal embedding in the NP: attributive adjectives	
	Nonfinite complement clauses controlled by common verbs (esp. <i>want</i>)		
3	Nonfinite complement clauses controlled by a wider set of verbs	<i>That</i> relative clauses, esp. with animate head nouns	
		Simple phrasal embedding in the NP: nouns as pre/postmodifiers	
		<i>Of</i> phrases as postmodifiers	
		Simple PPs as postmodifiers, esp. with prepositions other than <i>of</i> (concrete, locative meanings)	
4		Nonfinite relative clauses	
		More phrasal embedding in the NP (attributive adjectives, nouns as premodifiers)	
		Simple PPs as postmodifiers, esp. with prepositions other than <i>of</i> (abstract meanings)	
5	Preposition + nonfinite complement clause	Extensive phrasal embedding in the NP (multiple PPs as postmodifiers, with levels of embedding)	Appositive noun phrases

Table 10: Developmental stages for complexity features pertaining to the realisation forms of the object (adapted from Biber et al. 2011: 30-31)

The primary focus in clausal realisations is on predicates, where we can expect variance in frequency among the three studied groups. Phrasal realisations are concentrated around different modification types within the phrases, which can be phrasal or clausal (as shown in 5.5.1). In this case, the main focus is on embedding and hypotaxis. The last dimension, appositive and coordinative realisations, focuses on relationships on the same level, parataxis.

5.5.1 Clausal realisations of objects

In connection with objects, there are four clausal grammatical structures (or structural types) that may realise objects: (a) finite complement clauses controlled by common verbs (e.g. *think, know, say*) (stage 1)²⁶; (b) finite complement clauses controlled by a wider set of verbs (stage 2); (c) nonfinite complement clauses controlled by common verbs (stage 2); and (d) nonfinite complement clauses controlled by a wider set of verbs (stage 3).

Firstly, finite clauses will be analysed²⁷. As has already been stated, finite clauses as object realisations occurred in the L1-expert group the most; however, the differences in the number of finite clauses between the groups are insignificant. Significant differences arise when main clause verbs (or controlling verbs) are analysed. Table 11 and Figure 3 show that half of the verbs used by L2-novice writers are very common (their frequency rank, according to *new-GSL*, is below 147). Some examples of these common verbs are *know, say, see, show*. As for L1-novice writers, the median of their ranks is 297, thus, higher than that of L2-novice writers. Regarding the L1-expert group, half of the controlling verbs used with finite clauses have a frequency rank higher than 531, including the verbs *note, argue, propose, remark* or *urge*. To relate these findings to the developmental stages, it can be asserted that the structures produced by L2-novice writers correspond with the 1st developmental stage, whereas both groups of L1 writers correlate with higher developmental stages (stage 2 and above).

L1-expert			L1-novice			L2-novice		
verb	rank	11	verb	rank	8	verb	rank	10
argue	780	1	believe	237	2	bring	167	1
learn	331	1	confess	2500	1	know	61	1
note	730	1	examine	956	1	say	32	3
play	230	1	mean	165	1	see	57	1
propose	1037	1	reveal	818	1	show	126	1
remark	2500	1	see	57	1	showcase	2500	1
say	32	2	suggest	297	1	suggest	297	1
see	57	1				wonder	839	1
think	91	1						
urge	1854	1						

²⁶ The terms for the grammatical structures with their stages are adapted from Biber et al. (2011).

²⁷ In the studied sample, there was one special realisation of an object – a finite clause coordinated with a noun phrase, and this realisation will be identified together among other realisations by finite clauses.

Table 11: Controlling verbs of finite complement clauses in the 150 randomly chosen samples

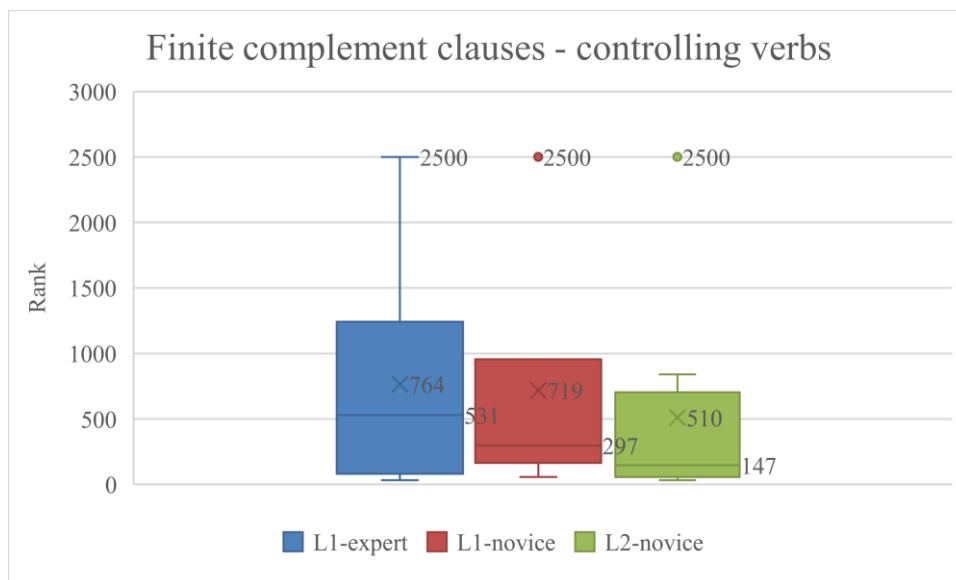


Figure 3: The verb rank of controlling verbs of finite complement clauses according to *new-GSL* in all three subcorpora²⁸

Secondly, complementation by non-finite clauses will be discussed (cf. Table 12, Figure 4). Since our sample consists of only 150 sentences, there were only 12 cases of non-finite clauses realising objects; thus, it is challenging to determine some tendencies. Nevertheless, it is obvious that L2-novice writers are faced with the linguistic challenge of not having the target language as their mother tongue. This claim may be supported by the fact that there are two overlapping controlling verbs (*attempt*, *seek*) between L1-expert and L1-novice writers, and the frequency rank of both of these verbs is relatively high. In addition, it is interesting that in both L1 groups, the medians increased (in comparison with finite clauses), which may imply that L1 writers tend to use less frequent controlling verbs with non-finite clauses compared to those verbs used with finite clauses. What is more, the increase in the L1-novice group was quite substantial (finite clauses (median 297) vs. non-finite clauses (median 522)). The only two controlling verbs complemented by non-finite clauses by L2-

²⁸ The figures have been rounded to whole numbers.

novice speakers have very different ranks, and this points to the fact that more data is needed to draw any conclusions.

L1-expert			L1-novice			L2-novice		
verb	rank	5	verb	rank	5	verb	rank	2
attempt	1182	1	attempt	1182	1	intend	901	1
exhort	2500	1	continue	298	1	know	61	1
fail	556	1	manage	545	1			
help	185	1	mean	165	1			
seek	522	1	seek	522	1			

Table 12: Controlling verbs of non-finite complement clauses in the sample

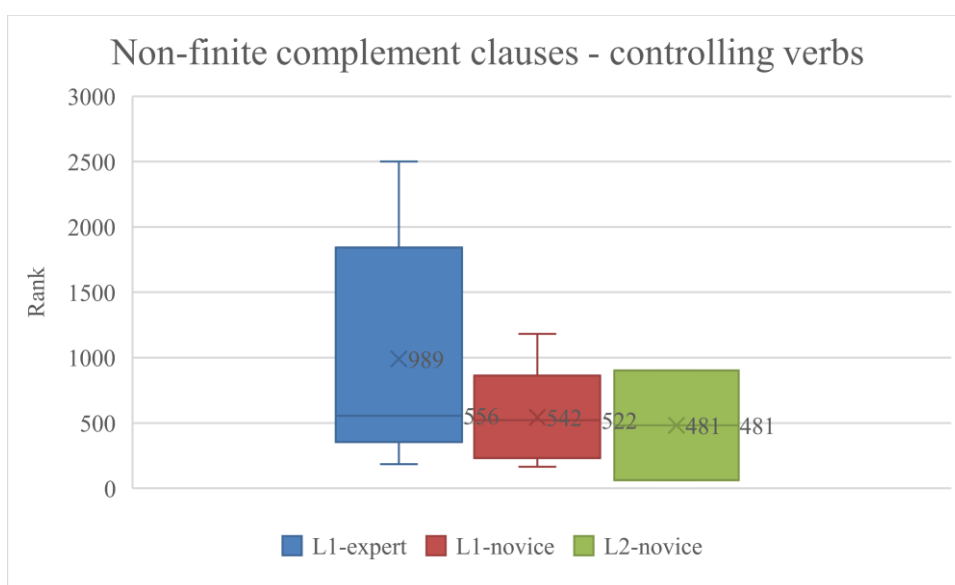


Figure 4: The verb rank of controlling verbs of non-finite complement clauses according to *new-GSL* in all three subcorpora

5.5.2 Phrasal realisations of objects – noun phrases

As already mentioned in the previous parts of this thesis, complex NPs form a majority of noun phrases in the studied sample. In L1-expert texts, 75 % of examples include complex noun phrases; in L1-novice texts, complex NPs account for 43 %; and in L2-novice texts, the number is higher as there are 57 % of complex noun phrases.

Firstly, there were 21 complex NPs in L1-expert texts, and in ten of these examples, the complexity of the noun phrase resides in embedding on the second level; most commonly, the complexity of the phrase is achieved by the combination of premodification by one or two adjectives and postmodification by a prepositional phrase (ex. 18), combining stages 2

and 3. Other sources of complexity on the second level only (i.e. there are no further levels of embedding) are (a) postmodification by a PP, mostly with the preposition *of* (stage 3, ex. 19) or (b) premodification by an adjective or a noun phrase (stage 3, ex. 20).

(18) Again , Cowley enhances [[**the stylistic and thematic**] **disarray** [**of Pindar's odes**]] [...]. (L1-expert)

(19) Despite these concerns, however, alehouse regulation was scant and attempts at enforcement left [**many records** [**of overt resistance**]] to the law. (L1-expert)

(20) Resistance theory only occasionally penetrated [**the** [**'good wife'**] **ideology**] [...]. (L1-expert)

The complexity of the rest of the examples is based on embedding on more levels than just on the second one. In the majority, there are three levels of embedding, combining the same complexity devices as when there are only two levels of embedding. There are several cases of multiple postmodification on different levels (stage 5, ex. 21) and one example of postmodification by coordinated noun phrases (stage 5, ex. 22). There was one example which displayed five levels of embedding, including postmodification by PP (stage 3), postmodification by a finite clause (stage 5), postmodification by a non-finite clause (stage 5) and postmodification by a relative non-finite clause (stage 5) (ex. 23). All these features attest to higher levels of structural complexity of academic papers written by expert native speakers.

(21) Nowhere else in the play do we witness [**anything** [**like this display** [**of skilled labor**] [**from Volpone**]] [...]. (L1-expert)

(22) In recent years critics have highlighted [**the** [**violent**] **subtexts** [**of Shakespeare's comedy**]: [[**the ritu- ahtic punishments of Falstaff**], [**the quarrels of the Windsor locals**], and [**the unhappy marriage of Alice and Frank Ford**]]]. (L1-expert)

(23) Cynthia Herrup also notes [**differences** [**in the rate** [**in which execution was ordered** [**depending upon both the type and value of goods** [**stolen among charges more serious than theft.**]]]]]] (L1-expert)

Secondly, L1-novice texts will be related to hypothesised developmental stages. In the studied sample, there were 12 complex NPs and 16 simple NPs, which is rather interesting

as this is the only group with more simple noun phrases than complex ones. A possible explanation for this phenomenon is that Czech students of English may seek to emphasize modifying noun phrases (i.e. overuse modification) because of the instructions they are being given during their training. Out of the twelve complex NPs, the complexity of nine was achieved on the second level of embedding. The most common complexity device was a combination of premodification by one or two adjectives (or a genitive in ex. 24) with postmodification by a prepositional phrase (ex. 25), a combination of stages 2 and 3, the same as in L1-expert texts. Some other devices displaying a higher degree of complexity of the complex NPs were based on postmodification by a prepositional phrase (stage 3, ex. 26), postmodification by a non-finite clause (stage 5, ex. 27) or premodification by an adjective combined with two coordinated postmodifying PPs (stage 2 and 5, ex. 28).

- (24) Her demonic capabilities are thoroughly compelling and exemplify **[a [woman's] capacity [for ingenuity.]]** (L1-novice)
- (25) Christina Rossetti's *The Iniquity of the Fathers Upon the Children* similarly defends **[a [disgraced] class [of people]]** [...]. (L1-novice)
- (26) Haemon intelligently begins **[a plea [to his father]]** [...]. (L1-novice)
- (27) He supports this also by referring to the candle carried by Othello which also suggests the "church' tableau' and sees **the imagery used by Shakespeare** as a possible hint to how the scene might be staged. (L1-novice)
- (28) Instead of reporting a single event (Earthmen meeting Mercurians for the first time), it involves **[a [wider] scale [in [time] and [space]]** [...]. (L1-novice)

The complexity of the three remaining examples took place on the third level of embedding; however, no specific tendencies can be observed as the complexity devices differed in the examples.

Lastly, some observations based on L2-novice texts in relation to the hypothesised developmental stages will be presented. As in L1-expert samples, the prominent noun phrase structural type was a complex NP (17 instances). In a majority of cases (13 instances), the source of complexity is performed on the 2nd level of embedding. The most usual complexity devices are premodification by an adjective (stage 2, ex. 29) or postmodification by a prepositional phrase, always with the preposition *of* (stage 3, ex. 30).

- (29) His name fits [**his [arrogant] personality.**] (L2-novice)
- (30) The poetic persona explores [**the possibilities [of their existence]**] [...]. (L2-novice)

In addition, there were four objects realised by a complex NP where there were more than two levels of embedded modifiers. The most complex of these was composed of 66 words and included five levels of embedding (ex. 31). There were also other devices contributing to a higher degree of complexity of the complex NP, namely, structures combining both premodifiers and postmodifiers (ex. 32) and coordinated clausal postmodifiers (ex. 33).

- (31) The clear and concise structure matches [**the argument [of the poem], [which is [introduced in the first stanza] and [elaborated upon in the following stanzas], [listing the manners [in which the lyrical subject has tried to vow the lady - “serving still”, “my goodwill” (stanza 1, lines 1 and 3), “all my smart” (stanza 3, line 2) as well as his “pain” (stanza 2, line 2) and which have all been ignored]]]]]. (L2-novice)**
- (32) Wyatt used **the so called rhyme royal , which is a pattern that goes in every of the three stanzas.** (L2-novice)
- (33) Hereby he supports **the notion that women are more deceitful and merely beings driven by their primal urges if not controlled by an outside party and need to be treated like property.** (L2-novice)

To sum up, relating the findings about complex noun phrases to the hypothesised developmental stages brings many surprising observations. Two major phenomena arose from the analysis: one is connected with the linguistic challenge (L1 vs. L2), and the other is connected with the academic challenge (novice vs. expert).

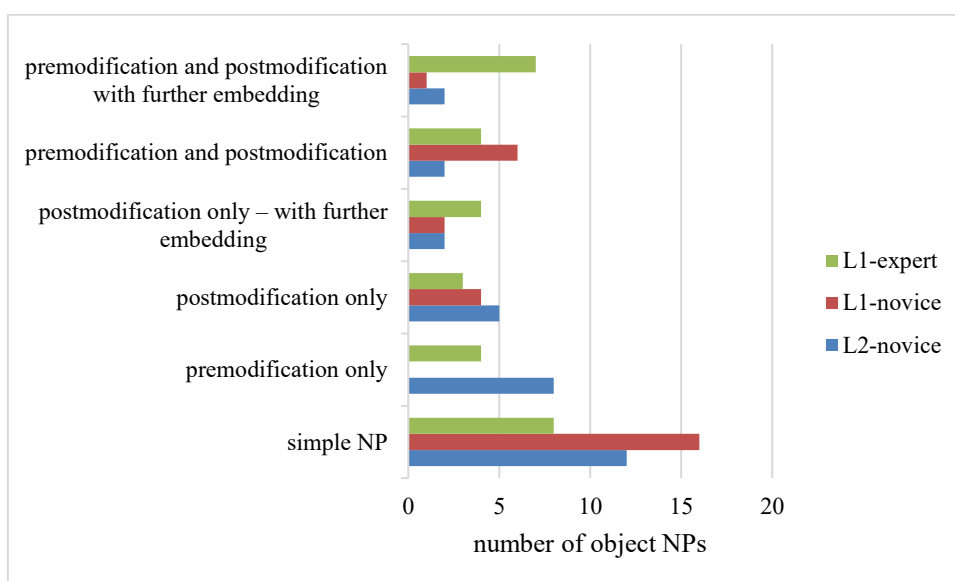


Figure 5: The complexity of objects realised by noun phrases (arranged in descending order of complexity, with the most complex NPs at the top of the vertical axis, and the least complex NPs at the bottom)

complexity feature	L2-novice	L1-novice	L1-expert
simple NP	12	16	8
premodification only	8	0	4
postmodification only	5	4	3
postmodification only – with further embedding	2	2	4
premodification and postmodification	2	6	4
premodification and postmodification with further embedding	2	1	7
total	31	29	30

Table 13: The complexity of objects realised by noun phrases (arranged in ascending order of complexity, with the least complex NPs at the top of the table, and the least complex NPs at the bottom)

As shown in Figure 5, objects realised by noun phrases with a less complex structure, i.e. simple NPs and NPs comprising merely modifiers or merely postmodifiers, are used more often by novice academic writers, both L2 and L1, than by expert L1 academic writers. Expert academic writers also use complex noun phrases with extensive phrasal embedding in the NP to a larger extent than novice academic writers. Since the extent of embedding in a noun phrase is considered an indicator of structural complexity, the differences between

expert writers, on the one hand, and novice writers, on the other, may reflect the ‘academic challenge’ that all novice academic writers have to face.

For L2 novice academic writers, it seems difficult to combine premodifiers and postmodifiers in a noun phrase. Compared to the L1 writers, both expert and novice, they overuse noun phrases which comprise merely premodifiers or merely postmodifiers. This is considered a feature typical of lower stages of structural complexity, which may be due to L2 writers’ English proficiency level.

In addition to the above quantitative differences, the three types of academic writers also differ in the range of realisation forms of modifiers within the noun phrase. The noun phrases used as objects by L2 novice writers are the most uniform ones in this respect, comprising mostly adjectival premodifiers and postmodifiers realised by *of*-prepositional phrases. While prepositional phrases constitute the most frequent realisation form of postmodification across all types of writers, L1 writers rely on a broader range of prepositions, which is considered a higher stage of complexity. Premodification by an embedded noun phrase was only attested in the L1-expert data.

5.5.3 Phrasal realisations of objects – prepositional phrases

Prepositional phrases as realisations of objects occur exclusively with prepositional objects. Thus, this chapter will describe the complexity of these objects, as demonstrated by the varied types of complex noun phrases realising them. Again, complex NPs constitute a majority of structural types realising prepositional objects: 67 % in L1-expert and L1-novice texts and 100 % in the L2-novice sample. It needs to be pointed out that due to our sample consisting of only 150 sentences, the number of prepositional phrases (and prepositional objects) is limited (22 instances); however, some tendencies can still be observed. Another very important thing to mention is that the structural type of complex prepositional phrases is composed of a preposition and a complex noun phrase on the first level of embedding.

Concerning the sample of L1-expert texts, there were only 4 complex prepositional phrases in total, and this number is the lowest of the three groups, which is connected with the fact that L1-expert writers employed prepositional objects the least. In two examples, the complexity is enhanced by premodification by two adjectives (ex. 34). The remaining two examples are more complex, one of them comprising embedding on four levels (ex. 35),

including the premodification by an adverb, postmodification by a PP and postmodifications by non-finite clauses. The other example features the coordination of four complex noun phrases (ex. 36). These combinations can be considered features of high stages of structural complexity.

- (34) In the public court, power derives [**from [professional and official] identity**]. (; L1-expert)
- (35) In so doing, she relies [**on [precisely the sort of ingenuity [exhibited by Susan Baskerville's cohort [in pursuing equitable remedies for the laws of coverture.]]]] (L1-expert)**
- (36) Moreover , short hand reporting accounts [**for [the differences in speech prefixes and names], [widely diverging lineations of verse passages], [instances of aural mistakes], and [variant stage directions.]**] (L1-expert)

As already mentioned, complex NPs complementing the preposition in prepositional phrases in L1-novice texts account for two-thirds of all of the prepositional phrases. One of the tendencies here is postmodification by a non-finite clause on the 2nd level of embedding (stage 5, ex. 37). In addition, L1 novice writers often use the combination of premodification by an adjective (or a genitive) with postmodification by a prepositional phrase as sources of complexity (ex. 38). There was one example with a higher degree of complexity, combining premodification by an adjective with postmodification by a finite clause on the second level, and postmodification by a non-finite adverbial clause on the third level (stage 5, ex. 39).

- (37) The protagonist looks [**for a way [to grasp reality and truth.]**] (L1-novice)
- (38) In a society where motherhood was synonymous with being a woman, Eulalie completely divorces herself [**from [society's] expectations [of womanhood] [...]**]. (L1-novice)
- (39) This stanza opens [**with three [simple] clauses [which supplement the undecorated diction of the child, [lacking the complexity of adverbs and adjectives.]**]] (L1-novice)

Concerning prepositional objects, L2-novice writers choose similar devices to enhance complexity as they do with direct and indirect objects. More specifically, combining premodification with postmodification of any kind is not frequent. Again, the pattern is to

use merely adjectives as premodifiers or merely prepositional phrases as postmodifiers on the second level (ex. 40a, b). Out of the seven examples featuring prepositional objects, only one displayed a higher degree of complexity with embedding on three levels in total (ex. 41), including postmodification by a finite adverbial clause on the second level (stage 2) and postmodification by two coordinated non-finite clauses on the third level (stage 5).

- (40) a. These three sonnets all deal **[with the [similar] topics.]** (L2-novice)
 b. Overall , while Petrarch , Spenser , and Sidney all concern themselves **[with the topic [of love]] [...]**. (L2-novice)
- (41) Poetry provided W . B . Yeats and Anna Akhmatova **[with a platform [where they were able [to employ their extraordinary talent - to create art with an impact.]]]** (L2-novice)

5.5.4 Coordination and apposition

This chapter looks at the paratactic dimension of structural complexity, analysing coordination and apposition.

Apart from embedding, structural complexity can also be enhanced by the use of multiple objects, complements or modifiers. Table 14 gives the numbers and realisation forms of multiple objects, but there were also further instances of coordination and apposition at lower levels of embedding. It is perhaps interesting that the number of multiple objects is the same in the L1 expert and L2 novice samples. The total numbers, however, are too low to allow any conclusions.

coordinated forms	L1-expert	L1-novice	L2-novice	total
clause + clause	3	2	3	8
NP + clause	1	X	2	3
NP + NP(s)	3	1	2	6
total	7	3	7	17

Table 14: Realisation forms of multiple objects

The complexity achieved by paratactic relations may be illustrated by ex. 42, where the prepositional object is realised by a *for*-prepositional phrase with a prepositional complement comprising a conjoin of four noun phrases. Multiple objects with appositionive relations among the constituents (ex. 43) were attested in the writing of all three types of writers. This suggests that apposition may be considered a high-stage complexity feature

typical of academic writing, which novice writers acquire at a relatively early stage of their academic career.

(42) Moreover, short hand reporting accounts [**for [the differences in speech prefixes and names], [widely diverging lineations of verse passages], [instances of aural mistakes], [and variant stage directions]**]. (L1-expert)

(43) Later on, he finally brings Volpone **what is supposed to be considered his most precious possession: Celia.** (L2-novice)

6 Conclusions

The analytical part of this thesis is divided into five major parts, each focusing on a different dimension or aspect of the structural complexity of objects. The contrastive interlanguage analysis approach adopted in this thesis has brought some interesting observations about the nature and level of structural complexity in texts produced by three different groups of writers – L1 experts, L1 novices and L2 novices. Moreover, the analysis has, in many different ways, confirmed the existence of two challenges that Czech students of academic English have to deal with – the academic and the linguistic challenges.

Firstly, the focus was on the analysis of the predicates of the main clause. These were examined from three points of view: their diversity, their lexical sophistication and their overlaps across the three studied groups.

Regarding the total number of all verb types used by each of the writer groups, the analysis has revealed that the diversity of verbs used by experts is slightly wider than that of novice writers, both L1 and L2. Thus, it is possible to state that novice writers are generally likely to use a narrower range of verbs than expert academic writers. Moreover, it has been proved that novice writers opt for the repetition of the same verbs more frequently than L1-expert writers.

For the analysis of the level of proficiency, or lexical sophistication, of the main clause verbs, the *New General Service List (new-GSL)* was used. After assigning the verbs to their ranks and counting the means of ranks for each of the groups, some remarkable findings arose. Generally, the means of the ranks gradually increased from L2-novice, through L1-novice to L1-expert texts, pointing to the fact that Czech learners use more basic verbs, L1 novices use slightly more advanced and less frequent verbs, whereas L1 experts display the highest level of sophistication of the used verbs. This proves that novice writers have to deal with the challenging nature of academic writing. In addition, it has been observed that there is a noticeable difference in the verb ranks of L2-novice and L1-novice groups, where the range of verb ranks is much wider in the L1-novice subcorpus.

The last area connected with the dimension of controlling verbs focused on overlapping verbs. A very striking observation is that novice writers, independent of whether English is

or is not their mother tongue, favour using more general and basic verbs such as *describe*, *have*, *lead*, *make*, *see*, *suggest*, *support* or *use*. When comparing either of the novice groups to L1 experts, a shared phenomenon here is the usage of verbs from the academic register (such as *present*, *provide*, *highlight* or *include*). Interestingly, only two verbs occurred in all three groups, namely, *make* and *see*. Both of these verbs are generally very frequent (according to *new-GSL*); thus, their employment in each of the subcorpora may be based on the fact that in appropriate situations, writers use very common verbs, independent of whether they are novice or expert writers.

Secondly, the analysis of the transitivity of the main clause predicates and the clause patterns has confirmed our expectations because the monotransitive pattern was the most frequently used clause pattern in our sample (appearing in 113 examples, 82 %). The analysis of ditransitive patterns has shown that writers from all three groups prefer using the pattern with the prepositional object over the one with the indirect one. A difference between L1 and L2 writers arose when the reversed order of the pattern with a prepositional object was examined, as L2 writers used this pattern only in its unmarked order (direct object preceding the indirect object). The complex transitive pattern has also brought one finding – both novice groups use complex transitive verbs much more frequently than L1 experts.

Thirdly, the length of the structures realising objects was computed, showing that expert writers generally use more extended structures for objects. In addition, the range of the number of words used by novice groups was much narrower than that of experts, which may imply that novice writers tend to use similar structures in terms of objects. As expected, the most frequently used structures in all groups consisted of single-word phrases. A difference, however, arose in the use of two-word and three-word structures. The two-word structure displaying a lower degree of structural complexity was used only once by expert writers, but novice writers used it quite extensively, confirming expert writers' tendency to use longer structures to realise objects in general. As for the three-word structures, the observation again points to the linguistic challenge of L2-novice writers since they used structures consisting of three words quite frequently, whereas such structures were not common in either of the L1 groups. Moreover, this shows that L2-novice writers have a tendency to employ premodification by adjective as the sole source of structural complexity.

Fourthly, the analysis of types of objects was held. As expected, direct objects represented a vast majority of our sample, and several structural types were used to realise them. The distribution of noun phrases was similar in all three samples. Finite and non-finite clauses were the structural types that revealed the differences between the studied groups. Finite clauses can be considered an inherently complex structural type, and this was confirmed by the fact that in the majority, they were used by L1 experts, pointing, again, to the challenge novice learners have to deal with if they want to become members of the academic world. Non-finite clauses uncovered evidence of the linguistic challenge as their use among L2-novice texts was sparse (only two instances), probably due to the impact of their mother tongue (Czech), where the frequency of use of non-finite clauses is lower than in English.

Additionally, the sources of complexity of all structural types realising direct objects were examined. Concerning noun phrases, complex noun phrases constituted a majority; however, novice writers used simple NPs more frequently than experts. One of the most striking findings of our analysis is that when examining the levels of embedding of the objects, the postmodification within a complex NP is usually phrasal; however, the deeper the level, the higher the probability of clausal features to occur. This particular finding confirms Biber et al.'s (2011) claims of the phrasal nature of written language, but it also adds to the finding of the structures' embedding level. As for finite clauses, the sources of complexity were often not connected to multiple levels of embedding as these structures often included no embedded clauses. In the cases when object clauses comprised embedded clauses, the source of complexity usually resided in postmodification by a finite or a non-finite clause. The source of complexity of non-finite clauses most frequently consisted in using two coordinated non-finite clauses.

Due to our sample consisting of only 150 examples, the analysis of indirect objects has not brought any noteworthy findings apart from the fact that all indirect objects were realised by a simple noun phrase, not displaying a high degree of complexity.

Within the parameter of types of objects, prepositional objects were analysed as well. The structural types of such objects were always prepositional phrases, though the complements of the prepositions, both clausal and phrasal, were varied and displayed different degrees of complexity. In prepositional complements realised by noun phrases, several sources of

complexity were analysed, and the results show that a combination of premodification and postmodification occurs only in L1 texts, showing that L2-novice writers may struggle with using more complex structures and prefer using simpler ones.

Lastly, the analysis was performed in relation to hypothesised developmental stages of structural complexity, as proposed by Biber et al. (2011). The findings related to clausal realisations of objects show that L2-novice writers often employ structures that correspond to the first developmental stage, while both groups of L1 writers usually employ structures correlating with higher developmental stages, displaying a higher degree of complexity. Several important findings can be mentioned regarding the results of the analysis of noun phrases in connection to the developmental stages. The impact of the academic challenge in employing complex noun phrases is manifested in expert writers' frequently using complex noun phrases with many levels of embedding on one hand and novice writers using noun phrases with less complex structures on the other. Moreover, L2-novice writers have to deal with two more phenomena due to their not being native speakers of English. One of them is overusing complex noun phrases that involve premodification or postmodification as the only source of complexity and, at the same time, struggling with using a combination of premodifiers and postmodifiers in the noun phrase. The other phenomenon subsumed under the linguistic challenge is based on the differences in the range of realisation forms of modifiers within the noun phrase since L2-novice writers use mostly adjectives as premodifiers and *of*-prepositional phrases as postmodifiers, both of which correspond to lower developmental stages. In addition, the paratactic dimension of coordination and apposition, based on using multiple objects, complements or modifiers, was analysed. Interestingly, all three groups use appositive or coordinated structures to realise objects, which may indicate that despite apposition being a complexity feature of a high stage, it is acquired by novice writers relatively early in their academic careers.

We hope that this diploma thesis has contributed to understanding complexity as a multi-dimensional construct, highlighting the advantages of combining quantitative and qualitative research and Contrastive Interlanguage Analysis to describe the structural complexity of objects.

References

- "Automatic analysis of syntactic complexity in second language writing." *International Journal of Corpus Linguistics*, vol. 15, no. 4, 2010, pp. 474-496. John Benjamins Publishing Company, <https://doi.org/10.1075/ijcl.15.4.02lu>.
- Biber, Douglas, and Bethany Gray. "Challenging Stereotypes About Academic Writing: Complexity, Elaboration, Explicitness." *Journal of English for Academic Purposes*, vol. 9, issue 1, 2010, pp. 2-20. Elsevier B.V., <https://doi.org/10.1016/j.jeap.2010.01.001>.
- Biber, Douglas, and Susan Conrad. *Register, Genre, and Style*. 2nd edition, Cambridge: Cambridge University Press, 2019.
- Biber, Douglas, et al. "Should we Use Characteristics of Conversation to Measure Grammatical Complexity in L2 Writing Development?" *TESOL Quarterly*, vol. 45, no. 1, March 2011, pp. 5-35. Wiley Online Library, <https://doi.org/10.5054/tq.2011.244483>.
- Biber, Douglas, et al. "The linguistic organization of grammatical text complexity: comparing the empirical adequacy of theory-based models." *Corpus Linguistics and Linguistic Theory*, 15 June 2023. De Gruyter, <https://doi.org/10.1515/cllt-2023-0016>.
- Biber, Douglas, et al. *Grammar of Spoken and Written English*. Pearson Education Limited, 1999.
- Bulté, Bram, and Alex Housen. "Defining and operationalising L2 complexity." *Dimensions of L2 Performance and Proficiency. Complexity, Accuracy and Fluency in SLA.*, edited by Alex Housen, Folkert Kuiken and Ineke Vedder, John Benjamins Publishing Company, 2012, pp. 21-46.
- Callies, Marcus. "Learner Corpus Methodology." *The Cambridge Handbook of Learner Corpus Research*, edited by Sylviane Granger, Gaëtanelle Gilquin, and Fanny Meunier, Cambridge: Cambridge University Press, 2015, 35-56.
- "Complexity." *Oxford Advanced Learner's Dictionary*, Oxford University Press, 2024, <https://www.oxfordlearnersdictionaries.com/definition/english/complexity?q=complexity>. Accessed 5 April 2024.
- Dušková, Libuše, et al. *Mluvnice současné angličtiny na pozadí češtiny*. 4th ed., Academia, 2012.
- Gilquin, Gaëtanelle. "Learner Corpora." *A Practical Handbook of Corpus Linguistics*, edited by Magali Paquot and Stefan Thomas Gries, Springer, Cham, 2020, pp. 283-303.
- Granger, Sylviane. "Contrastive Interlanguage Analysis: A Reappraisal." *International Journal of Learner Corps Research*, vol. 1, issue 1, 2015, pp. 7-24. ResearchGate, <https://doi.org/10.1075/ijlcr.1.1.01gra>.

- Heuboeck, Alois, et al. "The BAWE Corpus Manual." *Coventry University*, Coventry University, June 2010, <https://www.coventry.ac.uk/globalassets/media/global/08-new-research-section/current-projects/bawemanual-v3.pdf>. Accessed 20 June 2024.
- Huddleston, Rodney, and Geoffrey K. Pullum. *The Cambridge Grammar of the English Language*. Cambridge University Press, 2002. *Cambridge Core*.
- Hunston, Susan. "Lexical Grammar." *The Cambridge Handbook of English Corpus Linguistics*, edited by Douglas Biber and Randi Reppen, Cambridge: Cambridge University Press, 2015, pp. 201-215.
- Hunt, Kellogg W. *Grammatical Structures Written at Three Grade Levels: NCTE Research Report No. 3*. ERIC, 1965. *ERIC Institute of Education Sciences*, <https://eric.ed.gov/>. PDF Download.
- Lu, Xiaofei. "Automated measurement of syntactic complexity in corpus-based L2 writing research and implications for writing assessment." *Language Testing*, vol. 34, no. 4, 2017, pp. 493-511. *Sage Journals*, <https://doi.org/10.1177/0265532217710675>.
- Malá, Markéta, and Veronika Raušová. "Academic and Linguistic Challenges: It-patterns in English Learner Academic Writing." *Pushing the Boundaries: Enhancing Non-Native Learners' Communicative Competence*, edited by Petra Huschová, Kateřina Keplová, and Helena Zitková, University of Pardubice, 2022, pp. 10-26.
- Malá, Markéta, et al. "VESPA_CZ: Learner corpus of written academic English by advanced L2 English university students, whose L1 is Czech." *Wiki Český Národní Korpus*, Ústav Českého národního korpusu FF UK, 21 December 2022, https://wiki.korpus.cz/doku.php/en:cnk:vespa_cz. Accessed 20 June 2024.
- Palotti, Gabrielle. "A simple view of linguistic complexity." *Second Language Research*, vol. 31, no. 1, 2014, pp. 1-18. *Sage Publications*, <https://doi.org/10.1177/0267658314536435>.
- Quirk, Randolph, et al. *A Comprehensive Grammar of the English Language*. Longman Group Limited, 1985.
- Smith, Sheldon. "The New General Service List (new-GSL)" *EAP Foundation.com*, EAP Foundation, 1 April 2024, <https://www.eapfoundation.com/vocab/general/newgsl/>, Accessed 2 July 2024.
- "The Varieties of English for Specific Purposes dAtabase (VESPA)." *UCLouvain*, Université de catolique de Louvain, <https://uclouvain.be/en/research-institutes/ilc/cecl/vespa.html>. Accessed 20 June 2024.

Sources

The L1-expert and L1-novice corpora were accessed at <http://www.sketchengine.eu> (last accessed 5 July 2024).

Adam Kilgarriff, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vít Suchomel. “The Sketch Engine: ten years on.” *Lexicography*, 1: 7-36, 2014.

“English Literary Renaissance.” *The University of Chicago Press Journals*, The University of Chicago Press Journals Division, <https://www.journals.uchicago.edu/journals/elr/about>. Accessed 20 June 2024.

Hilary Nesi, Sheena Gardner, Paul Thompson and Paul Wickens. *British Academic Written English (BAWE) corpus*. 2007, <https://www.sketchengine.eu/british-academic-written-english-corpus/>.

Markéta Malá, Gabriela Brůhová and Kateřina Vašků. *VESPA_CZ: Learner corpus of written academic English*, version 1, 21 Dec. 2022. Ústav Českého národního korpusu FF UK, Praha 2022. Available from WWW: <http://www.korpus.cz>.

“Renaissance Studies.” *JSTOR*, ITHAKA, <https://www.jstor.org/journal/renastudies>. Accessed 20 June 2024.

“Shakespeare Quarterly.” *Folger Shakespeare Library*, Folger Shakespeare Library, <https://www.folger.edu/research/shakespeare-quarterly/>. Accessed 20 June 2024.

Appendix 1 (L1-expert subcorpus: 50 examples analysed in the analytical part)

Number of hit	Text ID	Example
1	ELR13	Religious reformers sought to substitute the principles and practices of personal spirituality for the decadent forms of Laudian Anglicanism.
2	SQ08	The three lines perform a capsule survey of the English countryside, from its heights to its delimiting shores.
3	ELR09	The more he plays creator, in other words, the larger his project becomes and the more he insures that she not be contained by his handiwork.
4	ELR02	Nowhere else in the play do we witness anything like this display of skilled labor from Volpone , as Scoto Mantuano pitches his product, establishes its value, and promotes its universal physiological power, “To fortify the most indigest and crude stomach . . . for the vertigine in the head . . . the mal caduco, cramps, convulsions, paralyse, epilepsies, tremor cordial, retired nerves, ill vapors of the spleen, stoppings of the liver, the stone, the strangury, hernia ventosa, ilica passio” (2.5.94–103).
5	RS01	However , he only takes examples from the first three books of the poem , ignores Spenser's characteristic way of deliberately confounding the reader with the need for retrospective reading, and makes little mention of the Irish context.
6	RS01	However , he only takes examples from the first three books of the poem, ignores Spenser's characteristic way of deliberately confounding the reader with the need for retrospective reading , and makes little mention of the Irish context.
7	RS01	However , he only takes examples from the first three books of the poem, ignores Spenser's characteristic way of deliberately confounding the reader with the need for retrospective reading, and makes little mention of the Irish context.
8	SQ10	In both drama and life, speech implies audience and the contingency of response : Pisanio plays a crucial role here as interlocutor and witness, midwife to Innogens birthing of herself—a nativity that takes shape across several moments of time, from the givenness of her own physical parturition, to her self-disclosing election of Posthumus, to her demand here that her courage be acknowledged as a deed.
9	SQ10	In both drama and life, speech implies audience and the contingency of response: Pisanio plays a crucial role here as interlocutor and witness, midwife to Innogens birthing of herself—a nativity that takes shape across several moments of time, from the givenness of her own physical parturition, to her self-disclosing election of Posthumus, to her demand here that her courage be acknowledged as a deed.
10	RS04	In fact, Medea's sudden apprehension of their wedding chorus (Heroides 12. 143) forms an unsettling syncopation of the 'strong confused noyce' (138) of Spenser's own processional.
11	ELR02	Yet in his visceral aversion to Lady Politic's unsolicited presence, Volpone fails to register that he has in fact been operating according to the central tenet of her advice: mental well-being demands the pursuit of novelty.
12	SQ04	But it also disaggregates them , insofar as love (rather than mere desire) is shown to be incommensurable with a certain kind of subjection in which service slides through vassalage toward slavery.

13	ELR12	Even though he seems tinged with regret and cites "Too much liberty" (1.2. 124) as the cause of his imprisonment, this does not impress upon him a profound respect for authority in general, or Angelo in particular, Thus can the demi-god, authority, Make us pay down for our offence by weight.
14	ELR12	Even though he seems tinged with regret and cites "Too much liberty" (1.2. 124) as the cause of his imprisonment, this does not impress upon him a profound respect for authority in general, or Angelo in particular , Thus can the demi-god, authority, Make us pay down for our offence by weight.
15	ELR05	Reflecting on her words "forfeit" and "vantage," both of which can have commercial overtones, Ernst Leisi says, " The image represents Christ as a magnanimous creditor who renounces his profits although the souls are forfeited. "
16	RS08	Resistance theory only occasionally penetrated the 'good wife' ideology – Gouge exhorts wives to practice 'wisdom, patience, and obedience', even if a husband is an 'enemie of Christ', but eventually acknowledges the rare necessity of wifely defiance of a husband 'if her husband commands her to do that which God has expressly forbidden', he does not, however, include any cases of a father's violence towards the couple's children as one of the rare grounds for maternal intervention.
17	RS08	Resistance theory only occasionally penetrated the 'good wife' ideology – Gouge exhorts wives to practice 'wisdom, patience, and obedience', even if a husband is an 'enemie of Christ', but eventually acknowledges the rare necessity of wifely defiance of a husband 'if her husband commands her to do that which God has expressly forbidden', he does not, however, include any cases of a father's violence towards the couple's children as one of the rare grounds for maternal intervention.
18	RS08	Resistance theory only occasionally penetrated the 'good wife' ideology – Gouge exhorts wives to practice 'wisdom, patience, and obedience' , even if a husband is an 'enemie of Christ', but eventually acknowledges the rare necessity of wifely defiance of a husband 'if her husband commands her to do that which God has expressly forbidden', he does not, however, include any cases of a father's violence towards the couple's children as one of the rare grounds for maternal intervention.
19	RS08	Resistance theory only occasionally penetrated the 'good wife' ideology – Gouge exhorts wives to practice 'wisdom, patience, and obedience', even if a husband is an 'enemie of Christ', but eventually acknowledges the rare necessity of wifely defiance of a husband 'if her husband commands her to do that which God has expressly forbidden' , he does not, however, include any cases of a father's violence towards the couple's children as one of the rare grounds for maternal intervention.
20	RS08	Resistance theory only occasionally penetrated the 'good wife' ideology – Gouge exhorts wives to practice 'wisdom, patience, and obedience', even if a husband is an 'enemie of Christ', but eventually acknowledges the rare necessity of wifely defiance of a husband 'if her husband commands her to do that which God has expressly forbidden', he does not, however, include any cases of a father's violence towards the couple's children as one of the rare grounds for maternal intervention.
21	SQ10	For James and for Duncan-Jones, The Tempest, broached through a series of careful negations, yields nothing as certain as a self-portrait , but does permit a sustained encounter with the poet.
22	SQ10	For James and for Duncan-Jones, The Tempest, broached through a series of careful negations, yields nothing as certain as a self-portrait, but does permit a sustained encounter with the poet.

23	SQ03	In so doing, she relies on precisely the sort of ingenuity exhibited by Susan Baskerville's cohort in pursuing equitable remedies for the laws of coverture.
24	ELR12	Cynthia Herrup also notes differences in the rate in which execution was ordered depending upon both the type and value of goods stolen among charges more serious than theft.
25	RS05	Accordingly, Spenser's Amoretti XVI provides neat poetic rendition of this aspect of the 'glauncing sight' where 'legions of loues with little wings' fly in mimicry of the pneumatic beam, 'darting their deadly arrowes fyry bright,/ at every rash beholder passing by'.
26	SQ06	Erne thinks the cuts were made by Shakespeare's acting company, perhaps with Shakespeare among the redactors , but the many remaining textual variants present a puzzle: "Parts of [Q1's] verbal texture are simply not Shakespearean."
27	SQ06	Erne thinks the cuts were made by Shakespeare's acting company, perhaps with Shakespeare among the redactors, but the many remaining textual variants present a puzzle: "Parts of [Q1's] verbal texture are simply not Shakespearean."
28	ELR09	If we see him more clearly now that he has named himself and become a character, we can only watch him amid a crowd.
29	ELR12	Edward Coke also remarks in the Institutes that "If in a large field a man might see together all the Christians that but in one year, throughout England, come to that untimely and ignominious death, if there were any spark of grace or charity in him, it would make his heart bleed for pity and compassion."
30	SQ11	Later that same year, men conducting an inquiry into local malt production noted that they could not complete their survey with respect to Quiney "by reason he is now at London."
31	ELR01	In recent years critics have highlighted the violent subtexts of Shakespeare's comedy: the ritualistic punishments of Falstaff, the quarrels of the Windsor locals, and the unhappy marriage of Alice and Frank Ford.
32	ELR03	Protestants had attempted to "liberate" the Word of God from behind the barriers imposed by the Catholic Church: the priest's interpretive authority, the linguistic difficulties imposed by Latin, the wall of scholastic commentary, and the plethora of ancillary books and texts (breviaries, horae, psalters, missals, graduals) that circulated in the Catholic devotional universe.
33	ELR12	Despite these concerns, however, alehouse regulation was scant and attempts at enforcement left many records of overt resistance to the law.
34	ELR12	Despite these concerns, however, alehouse regulation was scant and attempts at enforcement left many records of overt resistance to the law.
35	RS09	In the tension between the subject's automatic recognition of proper proportion and the deliberate project of affective self-fashioning, Puttenham locates decorum : it is the virtue which links the bodily and the social, uniting them by emphasizing their equal investment in harmony and formal coherence.
36	SQ05	We belatedly learn that some aspects of Volpone's performance had a nonmimetic significance for him.
37	RS08	We might now see why Shakespeare did not create a Hermione more active and strong in opposing Leontes' tyrannous acts and rescuing her children by herself, and why he gave Paulina such a defining role in the play.

38	ELR01	William Carroll argues (ignoring local history) that the name Brainford "obviously suggests the witch's origin in Ford's brain," and that it "symbolize[s] the extremity of Ford's delusion."
39	ELR04	It publicly marks a past benefit rather than constituting a request for a future bene- faction.
40	ELR05	He says it is unlawful for a testator to make a legacy conditional on whether "he marrie according to the appointment, arbitrement or consent of some other" (notably it is the son's choice and inheritance that is in question), because that other party, usually the executor, "might make an hard choise for the . . . legatarie, either by reason of the dislike of the parties, inequality of age, disparitie of kinred, disagreeing in maners or such like," a list that reads like the causes why the course of true love never did run smooth.
41	SQ05	In the public court, power derives from professional and official identity.
42	SQ11	Having decided not to appeal to the lord of the manor herself, Elizabeth pro posed that Lady Greville should be the one to write her husband "concerning the £20 which she hopeth Sir Edward hath allowed you."
43	ELR05	Thus the Duke focuses on Shylock's "strange apparent cruelty" (4.1.21) and urges hopefully that "mercy and remorse" may impel Shylock not only to forgo the penalty but to "Forgive a moiety of the principal, / Glancing an eye of pity on his losses" (ll. 26◆27); the answer to his cruel excess in the forfeiture is to be a remorseful and sympathetic abandonment by him even of some of the money to which Shylock has undoubted right.
44	ELR05	Thus the Duke focuses on Shylock's "strange apparent cruelty" (4.1.21) and urges hopefully that "mercy and remorse" may impel Shylock not only to forgo the penalty but to "Forgive a moiety of the principal, / Glancing an eye of pity on his losses" (ll. 26◆27); the answer to his cruel excess in the forfeiture is to be a remorseful and sympathetic abandonment by him even of some of the money to which Shylock has undoubted right.
45	SQ06	Moreover , short hand reporting accounts for the differences in speech prefixes and names, widely diverging lineations of verse passages, instances of aural mistakes, and variant stage directions.
46	ELR13	Again , Cowley enhances the stylistic and thematic disarray of Pindar's odes to create a deliberately "enthusiastical" and "irrational" poetics designed to resist interpretive penetration.
47	RS04	Ovid's Heroides constitute here not a final source but a liminal interjection.
48	RS08	Cleaver favours childbeating , if necessary: 'Household chastisement is agreeable to God's will, [as] is evident out of the Proverbs.'
49	RS08	In line with the traditional 'good wife' ideology, they paint mothers as ineffective defenders of children.
50	SQ08	Such attention might also help to counterbalance the tendency always to return the early modern psyche to its material bases and objects with an admission that the importance of some of its parts lay in their capacity– potentially dangerous and potentially joyous–to foster immersions in wholly immaterial worlds.

Appendix 2 (L1-novice subcorpus: 50 examples analysed in the analytical part)

Number of hit	Text ID	Example
1	AH-3006c	The reversal of ideas about the seasons in the opening lines draws attention to this.
2	AH-3006c	The reversal of ideas about the seasons in the opening lines draws attention to this.
3	AH-6998e	Her demonic capabilities are thoroughly compelling and exemplify a woman's capacity for ingenuity.
4	AH-3006e	In the Elizabethan court it was the warrior and prolific writer Sir Philip Sidney that embodied this to perfection.
5	AH-6998b	This stanza opens with three simple clauses which supplement the undecorated diction of the child, lacking the complexity of adverbs and adjectives.
6	AH-3008i	To an extent we can see why Conrad is often deemed as being on the cusp of Modernism.
7	AH-3160a	Miranda seeks to get away from home , as she tells her family at the start of the film when they receive Nicholas' request for a companion for Katrine; her arrival at Dragonwyck and resulting relationship with Nicholas set off the story.
8	AH-3160a	Miranda seeks to get away from home, as she tells her family at the start of the film when they receive Nicholas' request for a companion for Katrine; her arrival at Dragonwyck and resulting relationship with Nicholas set off the story.
9	AH-3008c	Agard attempts to prove this is something positive to embrace and not be threatened by; 'I only armed wit mih human breath.' (1.30)
10	AH-3057b	Instead of reporting a single event (Earthmen meeting Mercurians for the first time), it involves a wider scale in time and space since it covers the whole life of the narrator.
11	AH-3012d	If Oliver Twist had made Cockney, Cockney could make my rewriting.
12	AH-6998e	In the opening scene of this play, she offends Miss Tesman by identifying her new hat as that belonging to the maid and later confesses to Brack that she consciously upset her for her own entertainment.
13	AH-6998e	In the opening scene of this play, she offends Miss Tesman by identifying her new hat as that belonging to the maid and later confesses to Brack that she consciously upset her for her own entertainment.
14	AH-6998e	In the opening scene of this play, she offends Miss Tesman by identifying her new hat as that belonging to the maid and later confesses to Brack that she consciously upset her for her own entertainment.
15	AH-3005e	In such poems as Strange Meeting he uses assonant endings (Dominic Hibberd 1981 p61): 'Now men will go content with what we spoiled.
16	AH-3001f	* Consequently the incident leads directly into a passage criticising slavery and celebrating English freedom.
17	AH-3008i	Preventing the interior monologue from advancing in a strictly chronological manner raises questions about the nature of progression; one of the novel's key tropes.
18	AH-3080b	These foolish characters 'flatter the reader ' by including one in the joke of poking fun at them, and are used critically to highlight the intelligence of other key characters, specifically Emma and Mr. Knightly.
19	AH-3110a	Later on, the young princes also manage to do this and to taunt Richard in III.

20	AH-6998a	Christina Rossetti's <i>The Iniquity of the Fathers Upon the Children</i> similarly defends a disgraced class of people by voicing Margaret, an illegitimate child who is interminably stained with the personal and social shame of her ignoble birth.
21	AH-3001a	In response to this, I will examine how the themes of gender relations, education and use of landscape are used as vehicles to write tragedy in George Eliot's <i>The Mill on the Floss</i> (1860) and Charles Dickens's <i>Great Expectations</i> (1861) , focusing specifically on the final sequences of each novel.
22	AH-3008d	Accepting Humm's belief that; 'gender is constructed through language and is visible in writing style;' using direct, pithy end-stopped sentences rarely coloured by emotion make Movern appear more masculine, as conventional male language is seen as ordered and rational.
23	AH-3007b	The dialogue is 'trying to free your mind', but it cannot do it for the viewer.
24	AH-3007b	The dialogue is 'trying to free your mind', but it cannot do it for the viewer.
25	AH-3129b	Dickens never reveals what Estella actually looks like , instead only supplying the reader with the vague and superficial, she is, 'beautiful and self possessed.'(Chp 8, pg 55).
26	AH-3157e	The term 'subversive' generally means to tackle conventions or to be revolutionary.
27	AH-3009a	Others include the rejection of the possibility of any absolute truth , because everything is relative to time, place, and the individual.
28	AH-3006d	In his poem 'I saw a Chapel all of Gold' Blake suggests that the psychology of religion and sexuality at this time is bound , giving a critique of the destructive ideological way that sexuality had been conceptualised.
29	AH-3008d	Movern 's inconsistent femininity leads me to agree with Anderson that; 'writing about women does not necessarily make a writer 'feminist.'
30	AH-6998a	In a society where motherhood was synonymous with being a woman, Eulalie completely divorces herself from society's expectations of womanhood, implying that these designs need to undergo drastic change before they can satisfy this social group.
31	AH-6998a	In a society where motherhood was synonymous with being a woman, Eulalie completely divorces herself from society's expectations of womanhood , implying that these designs need to undergo drastic change before they can satisfy this social group.
32	AH-3160d	The protagonist looks for a way to grasp reality and truth.
33	AH-3080d	He also uses enjambment in this extract, flowing the lines into each other in order to maintain the pace of the poem.
34	AH-3001f	In this opening quotation Edward Said highlights a sensitive and intimate topic of Romantic Literature.
35	AH-3012a	Swift 's hero bears several similarities with Defoe's.
36	AH-6998f	Euripides introduces the gods as brutal and vengeful figures in the opening of this play through Dionysus' explanation of his birth: </s> Dionysus; he who Semele of yore, 'Mid the dread midwifery of lightening fire, Bore, Cadmus' daughter.
37	AH-3008b	Phonologically , declarative syntactic structures create a colloquial, prosaic tone that is forfeited by using the past tense and enumerated instances of reported speech.
38	AH-3008f	Martin believes this moment makes Hawthorne and Hester one ; Hawthorne, N.
39	AH-3147b	Although his unstable narration means that he loses control of the narrative, this does not mean that narrative does not feature dominantly in the novel.

40	AH-3008j	Tennenhouse believes that 'Jacobean drama proves one thing..sexual relationships are always political.'
41	AH-3004b	He describes him as 'a fellow that many men liked', but in particular he himself.
42	AH-3006g	The fascinating construction of the consciousnesses of these characters gives rise to a vast number of questions about their identities, sexuality, and personal conflicts.
43	AH-6998f	Bacchus continues to prepare the audience for the tragic drama they should expect to unfold by emphasising his intentions in visiting Thebes: ...soon I will terribly show That I am born a god (2) <s> Bacchus' intentions are wholly destructive, directing a play which, 'rather than a cautionary tale, is a vision of total despair'.
44	AH-3012d	Liam , the young hero, has many problems facing him every day, including small time crime such as stealing cars and vandalism.
45	AH-3068a	The enjambment between stanzas three and four reflects the volta in the argument.
46	AH-3008j	Othello 's name disintegrates into a series of 'O''s ; 'O ill-starred wench!'"
47	AH-3006b	He supports this also by referring to the candle carried by Othello which also suggests the "church' tableau' and sees the imagery used by Shakespeare as a possible hint to how the scene might be staged.
48	AH-3006b	He supports this also by referring to the candle carried by Othello which also suggests the "church' tableau' and sees the imagery used by Shakespeare as a possible hint to how the scene might be staged.
49	AH-3160b	Their performances display their mutual affection which culminates in the dance.
50	AH-6998f	Haemon intelligently begins a plea to his father by feeding his ego with praise, declaring that he is subordinate to his father in everyway and he will never hesitate to obey his word.

Appendix 3 (L2-novice subcorpus: 50 examples analysed in the analytical part)

Number of hit	Text ID	Example
1	PRA0154-LIT-04	Poetry provided W . B . Yeats and Anna Akhmatova with a platform where they were able to employ their extraordinary talent - to create art with an impact .
2	PRA0154-LIT-04	Poetry provided W . B . Yeats and Anna Akhmatova with a platform where they were able to employ their extraordinary talent - to create art with an impact .
3	PRA0154-LIT-06	In other words , she regards marriage just one of the few occupations a woman is offered during Victorian times .
4	PRA0160-LIT-02	It is introduced to us already in the title of the story - the readers then know what to expect and look for , experiencing tension and anxiety from the very beginning , knowing that whatever happens , this story will not have a happy ending
5	PRA0055-LIT-05	The poetic persona explores the possibilities of their existence through their own mortality ; the speaker projects their life onto the horizon of their demise and becomes content with their " being-towards-death " .
6	PRA0055-LIT-05	The poetic persona explores the possibilities of their existence through their own mortality ; the speaker projects their life onto the horizon of their demise and becomes content with their " being-towards-death " .
7	PRA0001-LIT-01	This makes them appear so truthful in their blackness that everyone thinks them beautiful thanks to their sincerity and solidarity with the fallen beauty .
8	PRA0056-LIT-01	It is notable that when Shylock appears for the first time neither Bassanio nor the audience know what the Jew is thinking about - all that is clear is that he 's deep in contemplation
9	PRA0001-LIT-01	Since "to disgrace" could also signify "to make ugly", "to disfigure", it is interesting to let the poem's theme linger in mind and let words reflect the meaning: in this sonnet, "Art" has "disgraced" Nature by "disfiguring" her in a way, by making false effigies of true, natural beauty.
10	PRA0096-LIT-01	These three sonnets all deal with the similar topics.
11	PRA0157-LIT-09	Both authors show that the relationship between language structure and meaning is not logical and that language is an imperfect construct for communication .
12	PRA0160-LIT-02	The finale of the story contains the absolute extreme volume of emotions , as both Roderick and the narrator are scared to death , fear already mentioned as the most powerful of all emotions .
13	PRA0021-LIT-01	Sir John Davies 's " Sonnet 6 " from Gullinge Sonnets also rejects language of poetic conventions , albeit in a very different way .
14	PRA0057-LIT-01	As a result , the poet urges his lover in the last couplet to love him more than ever before and to enjoy the moments they have left .
15	PRA0092-LIT-01	Henry Howard , Earl of Surrey led a turbulent life , but was very disciplined in his verse .
16	PRA0055-LIT-05	The crux of Dickinson 's poetry is her unique ability to capture intricate ideas and concepts in the most austere manner ; this distinct tendency allows for various interpretations , depending on what meaning does the subjective reader seek to find in her lines .

17	PRA0074-LIT-01	Eve Sedwick says that : " Sparkish is disastrously candid about the purely instrumental , symbolic value that Alithea has for him . Pinchwife , on the other hand , is forced to psychotic extremes of concealment in his unsuccessful attempt to withdraw his wife from circulation . "
18	PRA0048-LIT-01	The problem is plain , the beauty of the addressee exceeds over the beauty of nature and not only that , nature always changes and even the greatest " s ummer 's day " can be destroyed due to nature 's mutability .
19	PRA0098-LIT-01	Overall , while Petrarch , Spenser , and Sidney all concern themselves with t he topic of love , all three employ the medium of sight in a specific manner , which allows their sonnets to present the same subject matter from different points of view .
20	PRA0098-LIT-01	Overall , while Petrarch , Spenser , and Sidney all concern themselves with t he topic of love , all three employ the medium of sight in a specific manner , which allows their sonnets to present the same subject matter from different points of view .
21	PRA0098-LIT-01	Overall , while Petrarch , Spenser , and Sidney all concern themselves with t he topic of love , all three employ the medium of sight in a specific manner , which allows their sonnets to present the same subject matter from differe nt points of view .
22	PRA0159-LIT-02	The clear and concise structure matches the argument of the poem, which is introduced in the first stanza and elaborated upon in the following stanzas, listing the manners in which the lyrical subject has tried to vow the lady - "serving still", "my goodwill" (stanza 1, lines 1 and 3), "all my smart" (stanza 3, line 2) as well as his "pain" (stanza 2, line 2) and which have all been ignored.
23	PRA0058-LIT-01	Every object has a counter object , every element a counter element and ev ery movement a countermovement .
24	PRA0058-LIT-01	Every object has a counter object , every element a counter element and ev ery movement a countermovement .
25	PRA0058-LIT-01	Every object has a counter object , every element a counter element and ever y movement a countermovement .
26	PRA0092-LIT-01	Both formats were implemented expertly by Surrey , but we can see that the second version of a sonnet , that he built serves his purposes and style o f progression through the poem more elegantly .
27	PRA0003-LIT-01	His name fits his arrogant personality .
28	PRA0160-LIT-01	Hereby he supports the notion that women are more deceitful and merely beings driven by their primal urges if not controlled by an outside part y and need to be treated like property .
29	PRA0154-LIT-06	In addition , Margaret 's pursuit of her intellectual interests only completes Margaret 's list of " all ' social crimes ' for women " (207) .
30	PRA0033-LIT-01	When he visits him for the first time in the play , he says he has brought Volpone a pearl ; then he says he has a diamond for him , too .
31	PRA0033-LIT-01	When he visits him for the first time in the play , he says he has brought Volpone a pearl ; then he says he has a diamond for him , too .
32	PRA0160-LIT-03	While De Monfort is being suffocated by his passions and emotions , which eventually leads to his ultimate fall , Joanna Baillie paints his older sister as the voice of reason , even regarded by some as " rationally superior " to her brother .
33	PRA0057-LIT-01	But the sonnet mostly looks only at one kind of love - - unrequited love .

34	PRA0042-LIT-01	For the involvement of Volpone's role, he is not able to fully enjoy the results of his effort, as there are plenty of interruptions forcing him to always be ready to play his role, thus he gets almost no time off to be himself; nevertheless, he appears enjoy the sole game of deception throughout the play.
35	PRA0161-LIT-03	The lovers build a single " room " into the whole world : " Let us possess o ne world , each hath one , and is one " (2 ; 14) .
36	PRA0042-LIT-01	Androgyno , for example , having a name deriving from " the state of having characteristics of having both sexes or of being of intermediate sex " , being one of the first to openly bring up the change of sex , however , does not int end doing so solely in order to create humorous situations ; he suggests t hat being of only one sex makes one " stale and forsaken " , bringing a pinch of dynamicity into the play.
37	PRA0042-LIT-01	Androgyno , for example , having a name deriving from " the state of having characteristics of having both sexes or of being of intermediate sex " , being one of the first to openly bring up the change of sex , however , does not int end doing so solely in order to create humorous situations ; he suggests that being of only one sex makes one " stale and forsaken " , bringing a pinch of dynamicity into the play.
38	PRA0031-LIT-01	Through the time , masks expanded and obtained various forms such as the classical face hidden behind the mask or cross-dressing .
39	PRA0037-LIT-01	He does not just play the role , he almost becomes the parson , which is simi lar to the trial scene in The Merchant of Venice when Portia is able to convi ncingly play a law clerk .
40	PRA0164-LIT-01	The fifth stanza describes mostly the human nature and that most people would prefer recieving love and not have to give anything in return .
41	PRA0033-LIT-01	Later on , he finally brings Volpone what is supposed to be considered his most precious possession : Celia .
42	PRA0033-LIT-01	Later on , he finally brings Volpone what is supposed to be considered his most precious possession : Celia .
43	PRA0074-LIT-01	Shylock himself exposes Antonio 's generosity saying that : " He lends out money gratis and brings down/The rate of usance here with us in Venice . " (I , iii , 41-2)
44	PRA0174-LIT-02	Eliot likens their whispering to " wind in dry grass , " quietly rattling , yet o f no real consequence .
45	PRA0174-LIT-02	Eliot likens their whispering to " wind in dry grass , " quietly rattling , ye t of no real consequence .
46	PRA0095-LIT-01	Edmund Spenser , the author of the introductory quotation , mentions perso nalis ed Mutability in many forms in his two " Cantos of Mutabilitae " .
47	PRA0174-LIT-02	Every time they went to sleep , the soldiers had to wonder , whether they w ould open their eyes in the morning .
48	PRA0064-LIT-01	Wyatt used the so called rhyme royal , which is a pattern that goes in eve ry of the three stanzas .
49	PRA0085-LIT-01	Horner 's acquaintances present themselves moral and virtuous but are easil y prone to cuckold their husbands which is obvious both to the audience and other characters with the only exception of their husbands themselves .
50	PRA0160-LIT-01	The female characters in the plays The Merchant of Venice by William Shak espeare and The Country Wife by William Wycherley showcase how these beliefs about women influenced them and how they reacted to them .