

**UNIVERZITA KARLOVA – FILOZOFICKÁ FAKULTA
ÚSTAV ANGLICKÉHO JAZYKA A DIDAKTIKY**

DIPLOMOVÁ PRÁCE

Lenka Vaňková

Testování anglických kolokací u českých studentů angličtiny

Testing English Collocations in Czech Learners of English

Praha, srpen 2012

Vedoucí diplomové práce:

Prof. PhDr. Aleš Klégr

Prohlašuji, že jsem diplomovou práci vypracoval/a samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

Souhlasím se zapůjčením diplomové práce ke studijním účelům.

I have no objections to the MA thesis being borrowed and used for study purposes.

V Praze dne 13. srpna 2012

Děkuji Prof. PhDr. Aleši Klégrovi za cenné připomínky, ochotu a odborné vedení práce.

Abstrakt

Tato diplomová práce se zabývá anglickými kolokacemi a jejich testováním u českých studentů angličtiny. Význam kolokací v jazyce a frazeologie jako lingvistické disciplíny byl objeven teprve nedávno, a proto se teoretická část nejprve zaměřuje na ustálená slovní spojení a frazeologii z obecného hlediska. Je zde stručně nastíněna historie a základní principy frazeologie a následně hlavní přístupy k této lingvistické disciplíně, další část je potom zaměřena na ustálená slovní spojení, zejména na jejich kategorizaci a důležité typologie. Hlavním bodem práce jsou kolokace, se zaměřením na hlavní přístupy ke kolokacím, na kritéria sloužící k jejich odlišení od ostatních typů ustálených slovních spojení a z nich plynoucí definice a klasifikace. Dále je také stručně shrnut dosavadní výzkum kolokací.

Analytická část (založená na studii S. Granger) zkoumá a popisuje výsledky testování anglických kolokací u českých studentů angličtiny. Hlavním cílem této části je zjistit, zda (nebo do jaké míry) výsledky testů potvrdí výsledky Granger a také prozkoumat míru úspěšnosti českých studentů jak u (pasivního) rozpoznávání, tak u (aktivního) užívání anglických kolokací.

Abstract:

The present study focuses on English collocations and their testing in Czech learners of English. Since the importance of collocations and phraseology as a linguistic discipline has not been recognised until recently, these concepts are at first introduced from the general point of view, the attention is given to the history of phraseology and its underlying principles as well as to the major approaches to it. Second, phraseological units are described from the point of view of their categorisation and of some influential typologies of these units. The main focus of interest is collocations, particularly major approaches to them as well as criteria commonly used to describe and delimit collocations from other types of prefabricated units, and their definition and classification. Previous research on collocations in learner English is also outlined.

The analytical part (based on Granger's (1998) study) analyzes the results of testing English collocations in Czech learners of English. The main focus is on whether these results confirm or deny the results of Granger's study, as well as on the learners' knowledge of and ability to use collocations, particularly on their (non)attaining the native "ideal."

Table of Contents

List of Figures.....	6
List of Tables	6
Abbreviations.....	7
Part I – Theoretical Background.....	8
1. Introduction.....	8
2. Prefabricated units: From the periphery to the core	10
2.1. Prefabricated units and Phraseology	10
2.1.1. The rise of phraseology as a linguistic discipline.....	10
2.1.2. Two approaches to phraseology	15
3. Prefabricated units: Towards the definition(s) of a collocation.....	17
3.1. Categories of word combinations.....	18
3.2. Word combinations: influential typologies	20
3.3. Granger & Paquot’s reconciling the two approaches.....	24
4. Collocations	26
4.1. Major approaches to collocations.....	27
4.1.1. Frequency-based approach	27
4.1.2. Phraseological approach.....	28
4.1.3. Other approaches	29
4.2. Criteria commonly used to describe and delimit collocations.....	30
4.3. The definition and classification of collocations in this study	32
4.3.1. Amplifiers.....	33
5. Previous research on collocations in learner English	34
5.1. Granger’ Contrastive Interlanguage Analysis	36
6. Material and Method.....	41
6.1. The Cloze Test.....	41
6.2. The Significance of Collocation Test.....	43
Part II – Data Analysis.....	45
7. Analysis – Introduction.....	45
7.1. English amplifiers: General overview	45
7.2. Results of the Cloze test	46
7.2.1. Overall results.....	46
7.2.2. Non-native speakers: (non)attaining the native “ideal”.....	52
7.2.2.1. Combinations exclusive to the NS / NNS database.....	57
7.2.2.2. Combinations common to both NS and NNS database.....	59
7.3. Results of the Significance of Collocation test.....	60

7.3.1. Overall results.....	60
7.3.2. Significant collocation.....	65
7.4. Pedagogical implications.....	71
8. Conclusion	73
References.....	80
Corpora	82
Dictionaries.....	82
České resumé	83
Appendix	90

List of Figures

Figure 1: Cowie's (1988) classification of word combinations	21
Figure 2: Cowie's (1981) phraseological continuum	21
Figure 3: Mel'čuk's (1998) typology	22
Figure 4: Burger's (1998) typology.....	23
Figure 5: Distributional categories	24
Figure 6: The phraseological spectrum	25

List of Tables

Table 1: Categories of referential phrasemes (extract).....	32
Table 2: Raw frequencies of amplifiers based on the <i>BNC</i>	46
Table 3: Overall results of the excerption	47
Table 4: Raw frequencies of amplifiers based on the NS and NNS data.....	47
Table 5: Individual amplifiers demonstrating statistically significant differences in occurrences.....	49
Table 6: Individual amplifiers demonstrating statistically significant differences in the number of NS and NNS combinations	51
Table 7: NS: Most frequent amplifiers.....	53
Table 8: NNS: Most frequent amplifiers	53
Table 9: Combinations exclusive to native- or non-native speakers and combinations common to both	57
Table 10: Possible combinations (<i>BNC</i>) vs. suggested combinations (NNS).....	61
Table 11: Most frequent responses (NS + NNS).....	63
Table 12: The number of significant collocations in the <i>BNC</i> and the NNS database.....	66
Table 13: Significant collocations: the <i>BNC</i> vs. the NNS database	67

Abbreviations

BNC	The British National Corpus
EFL	English as a foreign language
ESL	English as a second language
ICLE	International Corpus of Learner English
L1	first language
L2	second language
NNS	non-native speaker
NS	native speaker

Part I – Theoretical Background

1. Introduction

The present study focuses on English collocations and their testing in Czech learners of English. Even though the importance of collocations has not been recognised until recently, collocations in general represent one of the most important phenomena from the point of view of language and language teaching. Recent studies have demonstrated that collocations, as well as other prefabricated units, are not only a very common linguistic phenomenon in language processing and use but also that they play an essential role in both these areas and as such they are crucial from the point of view of language production and understanding. The aim of the study work is to provide comprehensive and detailed description of English collocations as well as various approaches to them, and to test English collocations in Czech learners of English. The testing will be based upon the study “Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae” by Sylviane Granger (1998) and the main focus of the testing will be on whether the results of collocation testing in Czech learners of English will confirm or deny the results of Granger’s study (as well as results emerging from other similar studies). In addition it is expected that the results will reveal some specifics of Czech learners of English connected with their use and knowledge of collocations.

In the theoretical part of this study, prefabricated units are introduced with the focus on the factors that have led to the recognition of the importance of phraseological units and their moving from the periphery to one of the most central areas of linguistic interest, and on the approaches to phraseology in general. The attention is also given to the categories of word combinations and to influential typologies delimiting different types of prefabricated units. Since the central point of interest of the present study is collocation, major approaches to this phenomenon are presented together with criteria commonly used to describe and delimit collocations from other types of prefabricated units. Furthermore, the previous research on collocations in learner English is outlined.

The analytical part is based on the comparison of native and non-native speakers’ use of collocations. In particular, the data extracted from the *British*

National Corpus are compared and contrasted with the results of two tests given to non-native speakers of English. The attention is given to the non-native speakers' knowledge of collocations, i.e. their ability to produce and use native-like combinations, in order to find out to what extent the combinations suggested by Czech learners of English correspond with or reflect native collocations, i.e. to what extent Czech learners are able to attain the "ideal."

2. Prefabricated units¹: From the periphery to the core

Generally speaking, the importance of collocations as well as other “prefabricated units” (see below) has not been recognised until recently, yet the role of these units is crucial not only in language use but also in language teaching. Ever since the prefabricated nature of a language was recognised and has become one of the most central areas of linguistic interest, there has been a number of approaches to collocations (and other prefabricated units) going hand in hand with a number of definitions and classifications of collocations according to the approach selected. In order to understand the importance of collocations in the language, a brief survey into the history of prefabricated units as well as basic description of the main approaches to this phenomenon is necessary.

2.1. Prefabricated units and Phraseology

2.1.1. The rise of phraseology as a linguistic discipline

Phraseology is a linguistic discipline dealing with the study of word combinations rather than single words (Granger & Meunier, 2008: ix), particularly the study of “the structure, meaning, and use of word combinations” (Cowie, 1998: 26). The underlying principle of phraseology is that a language is not a system which is based upon and built by a mere combining of grammatical rules and the lexis (in which a speaker has a wide range of choices available to him), but rather a system in which lexis and linguistic context, or more accurately co-text, are interconnected and thus inseparable entities, cf. as Ellis (2008: 1) points out, “words mean things in the context of other words.” Thus, it can be stated that most of our language acquisition is, or rather should be, based upon learning not just single words, but rather words and their contexts, or more particularly co-texts, i.e. word combinations or multi-word expressions, for these play crucial role in establishing or understanding the meaning of words as well as their use. These word combinations have been variously

¹ Actually, there are various terms used to refer to word combinations (see Section 2.1.1.). Nevertheless, for the purpose of the present study, the term “prefabricated units” has been chosen for it corresponds (in my opinion) most precisely with and reflects their nature.

called “prefabricated units, prefabs, phraseological units, phraseologisms, (lexical) chunks, multi-word units, or formulaic sequences (Nesselhauf, 2005: 1). In general, they consist of two or more words and they are “lexically and / or semantically fixed to a certain degree.” On the basis of numerous researches made in this area, it has indeed been revealed that “by far the largest part of the English speaker’s lexicon consists of complex lexical items” and most of these complex lexical items are semi-productive. In addition, corpus-based studies have regularly found that “most of naturally occurring language...consists of recurrent patterns, many of which are phraseological” (ibid. 1).

Though the importance of phraseology has increased significantly, it was for a long time considered tangential and unimportant from the point of view of linguistics (at least in the Western traditions of scholarship). Early signs of the importance of prefabricated units came with collocation, which “came to notice of a number of English teachers in Japan around 1930” (Sinclair, 2008: xv). Twenty years later, Firth, Fries, and Harris laid the foundations of phraseology as a linguistic discipline in English linguistic tradition (Ellis, 2008: 1). The building stone of what was later to develop into phraseology was the realization of the inseparability of lexis and linguistic context. Firth’s widely quoted claim that “you shall know a word by a company it keeps” (Firth, 1957: 11) was the basis for structuralist linguistics. It treats language as “a self-contained relational structure, whose elemental constructions derive their forms and functions from their distribution in texts and discourse” (Ellis, 2008: 1, 3) and sees structural patterning at all levels of a language. Fries (1952) made a distinction between lexical and structural meaning, with “structural meaning concerning the patterns relating a particular arrangement of form classes to particular structural meanings” (Ellis, 2008: 1). On the basis of this distinction, language acquisition is “the learning of an inventory of patterns as arrangements of words with their associated structural meanings.” Harris (1982, 1991) also considered form and information (i.e. grammar and semantics) inseparable. He developed a mathematical theory called Operator Grammar focusing on how language carries information, proposing that each human language should be seen as “a self-organizing system in which both the syntactic and semantic properties of a word are established purely in relation to other words” (Ellis, 2008: 2).

However, in the 1960s, structuralism was replaced by generative approaches. Chomsky (1965, 1981) refused construction-specific rules and

developed the so-called Principles-and-Parameters approach developing the general grammatical rules and principles of Universal Grammar (ibid. 3). In other words, grammar became “top-down and rule-governed, rather than bottom-up and emergent” (ibid. 3). This approach condemned phraseology and all related concepts, such as patterns, constructions, formulas etc., to the periphery of a language as they became “no longer interesting for such theories of syntax” (ibid. 3).

Fortunately, the revival of phraseology started during the 1980s and the 1990s and it was closely connected with the rise of three new linguistic disciplines, particularly Cognitive linguistics, Construction grammar, and corpus linguistics, all of which – generally speaking – rejected the underlying idea of Generative linguistics that syntactic categories and relations are universal. In contrast, these disciplines treated syntactic categories and relations as both language- and construction-specific and argue that constructions were in fact central to the grammar (Ellis, 2008: 4). **Cognitive linguistics** is in general described as a set of related approaches sharing several fundamental assumptions rather than a single theory (Gries, 2008: 12). The idea underlying these approaches is that a language (Ellis, 2008: 5):

draws on basic condition, on perception, attention allocation, memory and categorization and that it cannot be separated from these as a distinct, modularized, self-governed entity, that knowledge of a language is integrated with our general knowledge of the world, and that language use and language function interact with language structure.

All this is closely connected with the fact that phraseology interconnects words, grammar, semantics, and social usage. As Ellis (2008: 5) points out, phraseology indeed “resonates with a wide range of research areas within Cognitive linguistics.” Furthermore, it casts away the strict separation between lexicon and grammar suggested by generative approaches (Gries, 2008: 13).

The essential unit in Cognitive linguistic is a so-called symbolic unit. A unit (in general) is defined as (Langacker, 1987: 57):

a structure that a speaker has mastered quite thoroughly, to the extent that he can employ it in largely automatic fashion, without having to focus his attention specifically on its individual parts for their agreement...he has no need to reflect on how to put it together.

A symbolic unit is then “a pairing of a form and a meaning / function, i.e. a conventionalized association of phonological pole (i.e. a phonological structure) and a semantic / conceptual pole (i.e. a semantic / conceptual structure)” (Gries, 2008: 13). The most important aspect of symbolic units lies in the fact that a speaker does not have to, or more particularly does not, analyse the internal structure of symbolic units (and thus of prefabricated units as well). Thus, the definition of a symbolic unit is very similar to that of a prefabricated unit – in fact, prefabricated units actually represent one type of symbolic units in this approach. All this contributes to the fact that phraseology and cognitive linguistics are indeed “nearly maximally compatible” (ibid. 14).

The degree of compatibility between phraseology and **Construction grammar** is similar to that between phraseology and Cognitive linguistics since the main differences between these two disciplines and phraseology are largely terminological (Gries, 2008: 14). In general, the basic notion of Construction grammar is that “all grammatical phenomena can be understood as learned pairings of forms (from morphemes, words, idioms, to partially lexically filled and fully general phrasal patterns) and their associated semantic or discourse functions” (Ellis, 2008: 4). As Goldberg (2006: 18) points out, “the network of constructions captures our grammatical knowledge in toto, i.e. It’s constructions all the way down.”

As Gries (2008: 14) further explains, the central unit in Construction Grammar is a so-called construction and it is in fact analogous to the symbolic units in the previous paragraph. Goldberg’s definition of a construction is as follows: “C is a construction iff_{def} C is a form-meaning pair $\langle F_i, S_i \rangle$ such that some aspect of F_i or some aspect of S_i is not strictly predictable from C’s component parts or from other previously established constructions” (Goldberg, 1995: 4). In other words, the definition of a construction is analogous to that of a symbolic unit except for a construction (according to Goldberg’s definition) requires non-compositionality (Goldberg uses the term non-predictability) – which is not the case of symbolic units and thus of prefabricated units as well in Cognitive linguistics. (Interestingly, in her recent work, Goldberg has changed her attitude and non-compositionality is no longer required. The frequency of an expression is sufficient in order to gain the status of a construction (Gries 2008: 14). Langacker’s approach to Cognitive grammar corresponds with that of Goldberg except for he does not regard non-compositionality as a defining feature of a construction. Thus, it can be summarised that symbolic units (and thus prefabricated units as well) and constructions are more or less identical concepts, even though a construction was originally slightly more specific concept requiring one non-predictable element (ibid. 14). Again, it has been shown that there is a high degree of compatibility between phraseology and Construction grammar as well as between phraseology and Cognitive linguistics described in the previous paragraph.

Moreover, there is one crucially important aspect shared by both Cognitive linguistics and Construction grammar making the compatibility between these two and phraseology even more prominent, i.e. the importance of actual frequencies of occurrence in both theories. As Gries (2008: 15) points out, Langacker’s Cognitive Grammar is explicitly usage based in two senses. First, performance, i.e. exposure to

and use of, symbolic units is assumed to shape the linguistic system of both speakers and hearers. Second, sufficient frequency of occurrence is a necessary condition for a linguistic expression to gain the status of a unit. In addition, Goldberg's Construction Grammar also considers sufficient frequency of occurrence a necessary condition for a linguistic expression to gain the status of a construction (ibid. 15).

The third approach is probably most closely connected to phraseology for it was probably the most influential or powerful tool contributing to the revival of phraseology. **Corpus linguistics** and the rise of various corpora allowed analysis of large collections of written as well as spoken language. Results of various studies based upon corpus investigations have challenged many existing linguistic theories. Yet crucially, at least from the point of view of phraseology, corpus linguistic analyses have confirmed that natural language makes considerable use of recurrent words and constructions, thus the lexical context, or more particularly co-text, is crucial to knowledge of word meaning and grammatical role (Ellis, 2008: 4).

One of the central notions in corpus linguistics is a so-called pattern which can be defined as all the words and structures which are regularly associated with the word and contribute to its meaning. A pattern can be identified if a combination of words occurs relatively frequently, if it is dependent on a particular word choice, and if there is a clear meaning associated with it (Hunston & Francis, 2000: 37).

Again, there is a strong similarity and overlap between a pattern and a prefabricated unit as well as between a pattern and a symbolic unit and a construction respectively. This fact further contributes and supports the notion of phraseology as one of the "key concepts in both theoretical linguistics and in the method of corpus linguistics" (Gries, 2008: 17) in spite of different terminologies in each of the approaches.

Furthermore, since one of the main aspects of prefabricated units (and related concepts) is the frequency of occurrence, the corpus linguistics plays a prominent role from this point of view for it provides the frequency data which are obviously essential. More importantly, all these correspondences together with results of corpus investigation confirming such "distributional regularities" (Ellis, 2008: 4) have led to the formulation of one of the most prominent and crucial principles in contemporary corpus linguistics, i.e. Sinclair's idiom principle (Gries, 2008: 17) which should be applied as the first mode when analysing a text since most of text is can be interpreted on the basis of this principle (Ellis, 2008: 4). To put it briefly, according to this principle, a language is essentially "made up of strings of co-selected words that constitute single choices (Granger & Paquot, 2008: 29) and it sharply contrasts with the open-choice principle. Similar conclusion was reached for instance by

Kjellmer (1987: 140) who claims that “in all kinds of texts, collocations are indispensable elements with which our utterances are very largely made” or by Erman & Warren (2000) who estimated that “about half of the fluent native text is constructed according to the idiom principle” (in Ellis, 2008: 4-5). In addition, comparisons of spoken and written corpora have shown that prefabricated units are more frequent in spoken language (ibid, 5).

Thus, it has been showed that prefabricated units cannot be considered marginal. In contrast, they can be assumed to represent the core entities in both Cognitive linguistics and Construction grammar and their tremendous importance has been further demonstrated by researches in and subsequent results of corpus linguistics. Prefabricated units are not only very common phenomena in both spoken and written languages, but also, and more crucially, the phraseological research have also led to evidence for a claim by Pawley & Syder (1983: 213-215), who say that speakers’ mental lexicons do not contain only lexical primitives – in contrast, there are hundreds of thousands of phraseologisms (or prefabricated units) that could be “productively assembled but are, as a result of frequent encounter, redundantly stored and accessed. Thus, the analysis of phraseologisms does not only reveal patterns...of usage, but can also ultimately lead to more refined statements about matters of mental representation within the linguistic system” (in Gries, 2008: 18).

2.1.2. Two approaches to phraseology

As already mentioned above, phraseology is a linguistic discipline dealing with the structure, meaning, and use of prefabricated units. Prefabricated units come in many different shapes and forms, thus the scope of the field is the function of the criteria used by linguists to distinguish prefabricated units from non-prefabricated ones (Granger & Paquot, 2008: 27). From this perspective, there are generally two major approaches to phraseology (together with the delimiting of particular types of prefabricated units according to these approaches), i.e. the traditional, phraseological approach and the distributional, or frequency-based approach. The traditional approach has focused primarily on fairly fixed combinations (i.e. idioms, proverbs, etc.), whereas the more recent frequency-based approaches (based on corpus data) “have adopted a much wider perspective and included many word combinations that would traditionally be considered to fall outside the scope of phraseology” (ibid. 27).

As Granger & Paquot (2008: 28) point out, the Anglo-Saxon tradition has from the very beginning attached great importance to the less fixed category of collocation. In this section, an overview of the two major approaches to phraseology as well as the most important typologies will be presented in order to further clarify the scope of the field and the terminology used.

The traditional approach to phraseology in general originated in the former Soviet Union and other countries of Eastern Europe. This approach to phraseology (Granger & Paquot, 2008: 28):

restricts the scope of the field to a specific subset of linguistically defined multi-word units and sees phraseology as a continuum along which word combinations are situated, with the most opaque and fixed ones at one end and the most transparent and variable ones at the other. The core of this approach is considered idiomatic units, whose meanings cannot be derived from the meanings of the constituents.

Cowie's (1998) approach to phraseology is a direct descendant of this tradition. His continuum, i.e. free combinations || restricted collocations → figurative idioms → pure idioms (for details see Section 3.2.), directly follows early Russian schemes. One of the major concerns of linguists following this tradition has been to find criteria for distinguishing prefabricated units from one another as well as setting the boundary between one end of the scope – particularly the most-variable and transparent multi-word units – and free combinations, which fall outside the scope of phraseology for they are restricted only syntactically and semantically (ibid. 28). As Granger & Paquot (2008: 28-29) further point out, this approach “deserves much of the credit for having established phraseology as a discipline in its own right, created terminology for the field and provided linguists with a set of discrete criteria which can be used to categorize and analyze phraseological units.”

The second major approach to phraseology is closely connected with corpus linguistics and the frequency of co-occurrence. One of its main representatives is Sinclair who, instead of adopting “a top-down approach which identifies phraseological units on the basis of linguistic criteria, set up a bottom-up corpus driven approach to identify lexical co-occurrences” (Granger & Paquot, 2008: 29). As such, this approach works with a wide range of word combinations which do not all fit predefined linguistic categories which has led to opening up a “huge area of

syntagmatic prospection' encompassing sequences like frames, collocational frameworks, colligations, and largely compositional recurrent phrases" – all of which illustrate Sinclair's idiom principle. Thus, many of the units which were traditionally considered either peripheral or falling outside the interest of phraseology have now become central as it was revealed that they are pervasive in language whereas many of the most restricted units (representing the core of the phraseological approach) have proved to be highly infrequent (ibid. 29). Furthermore, Sinclair and his followers are not so much preoccupied with the delimiting of different linguistic categories / subcategories of word combinations or setting clearer boundaries to phraseology. In this tradition, phraseology is central from the point of view of a language: prefabricated units, "whatever their nature, take precedence over single words" (Granger & Paquot 2008: 29). Granger & Paquot (2008: 29-35) further point out that the two approaches to phraseology make its boundaries rather fuzzy as phraseology has to deal with everything, and so each approach relates to the four main areas of linguistics, semantics, morphology, syntax, and discourse differently.

3. Prefabricated units: Towards the definition(s) of a collocation

In the previous sections, we have seen that phraseologisms / prefabricated units are one of the core components of both written and spoken language. As already mentioned above, prefabricated units consist of two or more words and can be identified as recurrent, i.e. frequent, patterns, or multi-word expressions, in a language. Several important functions of prefabricated units can be identified. First, they play "an essential role in language learning, as they seem to be the basis for the development of creative language in first language and childhood second language acquisition" (Nesselhauf, 2005: 2). Secondly, in order to be fluent in both spoken and written language, the knowledge of prefabricated units is essential; as Nesselhauf (2005: 2) points out, following the results of psycholinguistic evidence, the human brain is "much better equipped for memorizing than for producing, and the availability of prefabricated units in human brain reduces processing effort and thus makes fluent language possible." Another important function of prefabricated units can be found in the fact that their use plays an important role in communicative process, i.e. it aids

comprehension in a way that “the recipient can understand the meaning of a passage of text without having to attend every word.” Importantly, whereas the use of natural, or native-like, prefabricated units indeed supports comprehension, non-natural, incorrectly used prefabricated units can “irritate the recipient and draw attention away from the message.” The last important function of prefabricated units lies in the fact that they “indicate membership of a certain linguistic group and fulfil ‘the desire to sound [and write] like others’” (ibid. 2).

So far, it is clear that the knowledge and correct use of prefabricated units is essential not only for native speakers, but also, and even more crucially, for learners of a language. It is generally agreed that prefabricated units have to be taught, though they are still “not treated adequately in English language teaching today” (Nesselhauf, 2005: 3). The situation is further complicated by the fact that there are many types of prefabricated units yet no unified approach to them, i.e. there is a large number of definitions of each type of prefabricated units according to the approach chosen. In the following sections, the most important types of prefabricated units are listed with subsequent commentary concerning various approaches to them.

3.1. Categories of word combinations

As already outlined in the previous sections, there are many differences between the typologies of word combinations or prefabricated units. These differences closely relate to the approach to phraseology chosen, in particular to the selection of features used to categorise multiword or prefabricated units and differentiate them from one another as well as the order of priorities given to particular features (Granger & Paquot, 2008: 34). Two (more or less) similar lists of features used to categorise these units and subsequently differentiate them from one another are presented below, Granger & Paquot’s (2008) and Gries’s (2008).

Granger & Paquot (2008: 35) give a list of five important features associated with multi-word units. As they claim, most classification favour one or more of these:

- i) internal structure (e.g. verb + noun or verb + preposition)
- ii) extent: phrase-level vs. sentence level
- iii) degree of semantic (non-compositionality)
- iv) degree of syntactic flexibility and collocability

v) discourse function

As they further point out, some categories (or subcategories) of multi-word units have acquired a relatively unified terms used to refer to them, for instance idioms being usually defined as non-compositional, whereas other categories are much more confusing from the point of view of terms used to refer to them, such as collocations which are used in a large number of different meanings (ibid. 35).

As a comparison, Gries's (2008: 4) so-called "notion of phraseology" used to delimit or specify particular types or categories of prefabricated units is also presented. Gries gives a list of six fundamental parameters that have to be taken into account in the analysis and differentiation of prefabricated units². It partially overlaps with Granger & Paquot's list, featuring some new parameters (iii. and iv.) but missing discourse function. Gries claims that it summarizes the set of parameters that are implied in the majority of studies concerning phraseological research and more importantly, it encompasses the two main approaches to collocations (for details see Section 4.1.):

- i. the *nature* of elements involved in a prefabricated unit;
- ii. the *number* of elements involved in a prefabricated unit;
- iii. the *number of times* an expression must be observed before it counts as a prefabricated unit;
- iv. the permissible *distance* between the elements involved in a prefabricated unit;
- v. the degree of *lexical and syntactic flexibility* of the elements involved;
- vi. the role that *semantic unity* and *semantic non-compositionality / non-predictability* play in the definition.

Gries also mentions additional or alternative criteria that can also be involved in the analysis of prefabricated units, i.e. possible separation of lexical flexibility and syntactic flexibility (or commutability / substitutability) of the elements involved in potential phraseological units and / or the distinction between encoding and decoding idioms (ibid. 4).

The first criterion, i.e. the nature of elements involved in a prefabricated unit, basically distinguishes prefabricated units into two groups, i.e. lexical units and grammatical ones. The third criterion, frequency of occurrence, is one of the most important criteria which has to be taken into account in order to classify a non-idiomatic word combination as a prefabricated unit. As Sinclair and his followers have argued, a word combination should be considered a

² In his list, Gries uses the term "phraseologisms."

prefabricated unit if observed frequencies exceed those expected on the basis of chance (ibid. 5). However, in majority of recent analyses, the frequencies or per centages of occurrence of prefabricated units have been restricted. The fourth criterion raises the question of whether a prefabricated unit can be discontinuous, i.e. containing some other elements in between, or the elements have to be in the closest proximity possible, i.e. right next to each other. Generally speaking, the usual measure of proximity between the elements of a prefabricated unit is a maximum of four words intervening (Sinclair, 1991: 170). The fifth criterion allows for completely inflexible patterns such as *by and large*; partially flexible patterns (e.g. *kick the bucket* disallowing passivisation); partially lexically-filled patterns; and patterns with completely unspecified lexically and thus maximally flexible expressions. Finally, the sixth criterion distinguishes between items functioning as semantic units, i.e. having “a sense just like a single morpheme or word,” and prefabricated sequences where non-compositional semantics is not a necessary condition.

To sum up Gries’s position, a prefabricated unit can be, in its widest sense, described as the “co-occurrence of a form or a lemma of a lexical item and one or more additional linguistic elements of various kinds which functions as one semantic unit in a clause or a sentence and whose frequency of co-occurrence is larger than expected on the basis of chance” (Gries, 2008: 6). Thus, in order to define a potential prefabricated unit, all these criteria or levels have to be taken into account. More importantly, once a word combination has been recognised as a prefabricated unit, these criteria subsequently help to recognize the particular type of the prefabricated unit in question.

3.2. Word combinations: influential typologies

Granger & Paquot (2008) also give some of the most important typologies of phraseological items to date which are useful as a starting point. The first typology to be mentioned is that of Cowie. It is one of the most influential typologies from the point of view of English lexicology and lexicography:

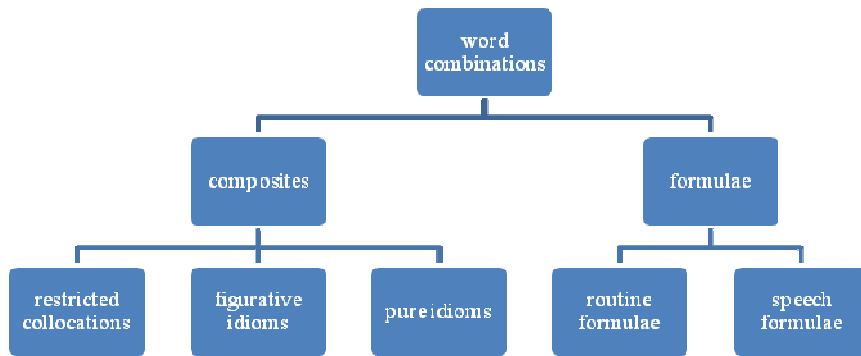


Figure 1: Cowie's (1998) classification of word combinations

As we can see in Figure 1 (adapted from Granger & Paquot, 2008: 36), the importance of this typology lies primarily in the fact that Cowie (1998) makes a primary distinction between *composites*, functioning syntactically at or below the sentence level, and *formulae*, which are pragmatically autonomous utterances. Composites are further subdivided into three categories, i.e. *restricted collocations*, *figurative idioms*, and *pure idioms* respectively. These categories represent a phraseological continuum constituted by the most transparent and variable multi-word units at one hand and the most opaque and fixed ones at the other:

free combination		restricted collocation	»	figurative idiom	»	pure idiom
<i>blow a trumpet</i>		<i>blow a fuse</i>		<i>blow your own trumpet</i>		<i>blow the gaff</i>

Figure 2: Cowie's (1981) phraseological continuum

Restricted collocations (often referred to simply as “collocations”) are characterised by restricted collocability and specialized or figurative meaning of one of the component. The difference between figurative and pure idioms respectively can be found in the fact that whereas the former have a figurative meaning yet also preserve a literal interpretation, the latter are semantically non-compositional and resist substitution of their components. The category of formulae consists of “sentence-like” units which “function pragmatically as sayings, catchphrases, and conversational formulae” (ibid. 36). Later, Cowie further subdivides this category into *routine formulae* which perform speech-act functions, such as *good morning*, or *see you soon*, and *speech formulae* used to “organize messages and indicate

speakers' or writers' attitudes," for instance *you know what I mean, are you with me?* etc. (ibid. 36).

Another influential typology is that suggested by Mel'čuk (1995, 1998; summarized in Granger & Paquot, 2008: 36-38) in the area of meaning-text theory. Mel'čuk's model is similar to Cowie's though the terminology is different:

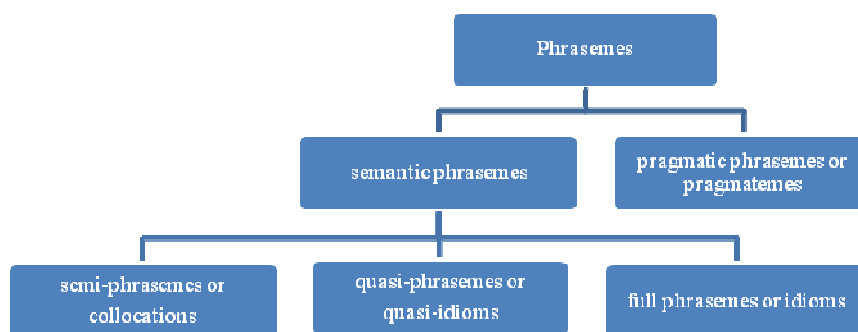


Figure 3: Mel'čuk's (1998) typology (in Granger & Paquot)

As shown in Figure 3, Mel'čuk's *semantic phrasemes* roughly correspond to Cowie's composites and his *pragmatic phrasemes* or *pragmatemes* are very close to what Cowie calls formulae, though he puts more emphasis on the functional, pragmatic aspect. More importantly, there is one crucial aspect of Mel'čuk's theory, i.e. his treatment of collocations by means of lexical functions. In other words, he attempts to describe lexical preferences – i.e. why it is possible to say *a heavy smoker* and *a big eater* whereas *a big smoker* and *a heavy eater* respectively does not sound natural in English – with lexical functions. According to Mel'čuk, a lexical function is “a very general and abstract meaning that can be expressed in a large variety of ways depending on the lexical unit to which this meaning applies.”

Examples of Mel'čuk's lexical functions are for instance:

- i) **Magn** which expresses the meaning of “intense(ly)” or “very” and functions as an intensifier, e.g. **Magn**(shave_N) = *close, clean*; **Magn**(easy) = *as pie, as 1-2-3*; **Magn**(to condemn) = *strongly*
- ii) **Oper** which expresses the meaning of “do / perform,” e.g. **Oper**₁(cry) = *to let out* [ART~]
- iii) **Real** which conveys the meaning of “fulfil the requirement of X” or “do with X what you are supposed to do with X,” e.g. **Real**₁(car) = *to drive* [ART~]; **Real**₁(accusation) = *to prove* [ART~]

The third example of an influential typology is that proposed by Burger (in Granger & Paquot, 2008: 37-38), who, unlike Cowie and even more than Mel'čuk, primarily concentrates on the function of phraseological units in discourse:

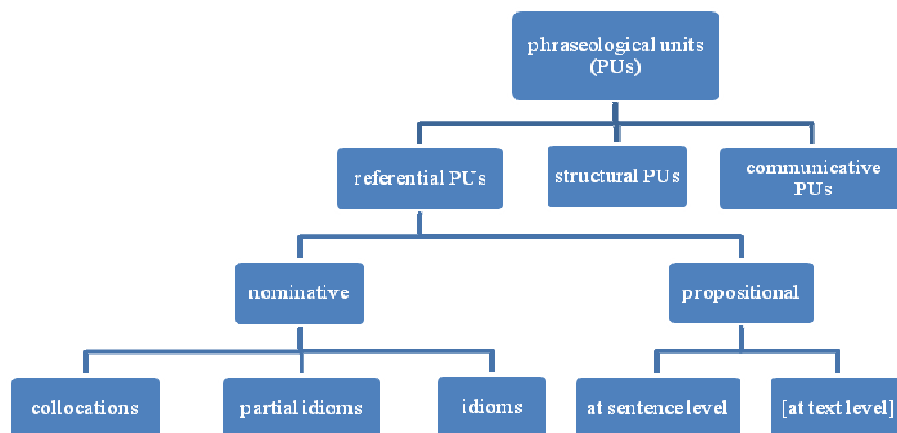


Figure 4: Burger's (1998) typology

Figure 4 shows that multi-word units, called *phraseological units* (PUs) by Burger, primarily divide into three main categories, i.e. *referential units*, *structural units*, and *communicative units*. Referential units are further subdivided into *nominal* and *propositional units* respectively on the basis of a syntactico-semantic criterion. The former category includes “constituents of the sentence and refers to objects, phenomena, or facts of life...and it broadly corresponds to Cowie’s composites.” In accordance with the Russian tradition and phraseologists such as Cowie and Mel'čuk, this category is again subdivided into *idioms*, *partial idioms*, and *idioms*. *Propositional* PUs in general function at sentence level (though they can also function at text level) and they refer to a “statement or an utterance about these objects or phenomena.” As such, this category includes proverbs and idiomatic sentences (corresponding with Cowie’s formulae and Mel'čuk’s pragmatic phrasemes). *Structural* PUs include constructions that establish grammatical relations, such as *as well... as...*, yet for Burger this category is the smallest and least interesting. In contrast, *communicative* PUs, or so-called *routine formulae* fulfil interactional function in the way that they are “typically used as text controllers to initiate, maintain and close conversation or to signal the attitude of the addressor,” for instance *Good morning*, or *Well, I mean...* .

3.3. Granger & Paquot’s reconciling the two approaches

Of the two approaches to phraseology outlined in 2.1.2., the emergence of the new, distributional one is, as Granger & Paquot (2008: 39-45) pertinently point out, proving to be of immense value to the field. However, both groups of linguists, i.e. those working in the traditional approach and those working with corpus-based methods of extractions and analysis, seem to be indifferent to or unaware of the benefits which the other approach offers. According to Granger & Paquot (2008: 41), any “rapprochement will only be fruitful if it is accompanied by some rigorous clarification of the terminology” and this is only possible with the clear distinction between the two typologies, i.e. to make one typology for the automated extraction and one typology for linguistic analyses.

Granger & Paquot (2008: 39) suggest that the terminology used in automated extraction should correspond with the so-called *distributional categories*:

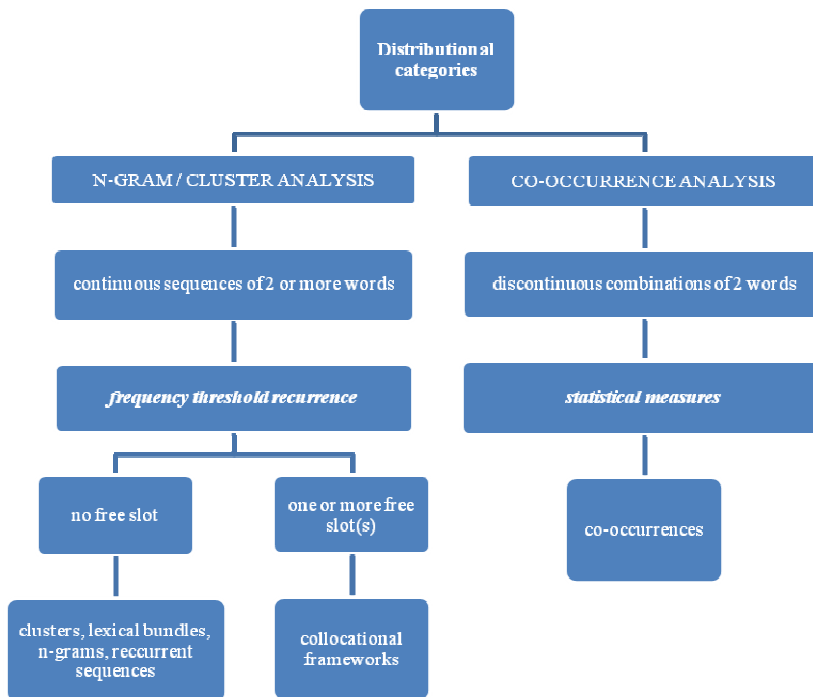


Figure 5: Distributional categories (Granger & Paquot, 2008)

Distributional categories emerged from the studies rooted in the distributional approach to phraseology and the primary distinction can be made according to two main extraction methods, i.e. n-gram analysis and co-occurrence analysis. N-gram analysis is a method allowing for the extraction of “recurrent continuous sequences of two or more words, viz. ‘recurrent expressions, regardless of their idiomacy, and

regardless of their structural status” (Biber et al., 1999: 990). There is a large number of terms used to refer to the extracted sequences, for instance n-grams, or more specific bigrams or trigrams (cf. Stubbs, 2007); clusters (Scott and Tribble, 2006); lexical bundles used by Biber; chains (Stubbs, 2002); recurrent sequences (De Cock, 2003); or recurrent word combinations (Altenberg, 1998). The examples of n-grams are sequences such as *Can I have a, in the case of, on the other hand, etc.* (Granger & Paquot, 2008: 39). In contrast, co-occurrence analysis is basically the “statistically uncovering of significant word co-occurrences.” The retrieved units are referred to as “collocations” or “collocates” (cf. collocational frameworks, the term connected with n-gram analysis, representing a special category of recurrent sequences, consisting of sequences of one or more free slots, such as *a + ? + of, be + ? + to, or too + ? + to*. Interestingly, the COBUILD project is built upon co-occurrence analysis and as Granger & Paquot (p. 40) further point out, co-occurrence analysis methods “constitute fantastic heuristic devices that show their full potential in a program like the *Sketch Engine*, which provides lexicographers with ‘corpus-based summaries of a word’s grammatical and collocational behaviour.’”

Concerning the linguistic classification, Granger & Paquot (2008) consider it essential to integrate the new insight derived from the corpus based approach. Accordingly they propose an extended function-oriented version of Burger’s classification (see Figure 6):

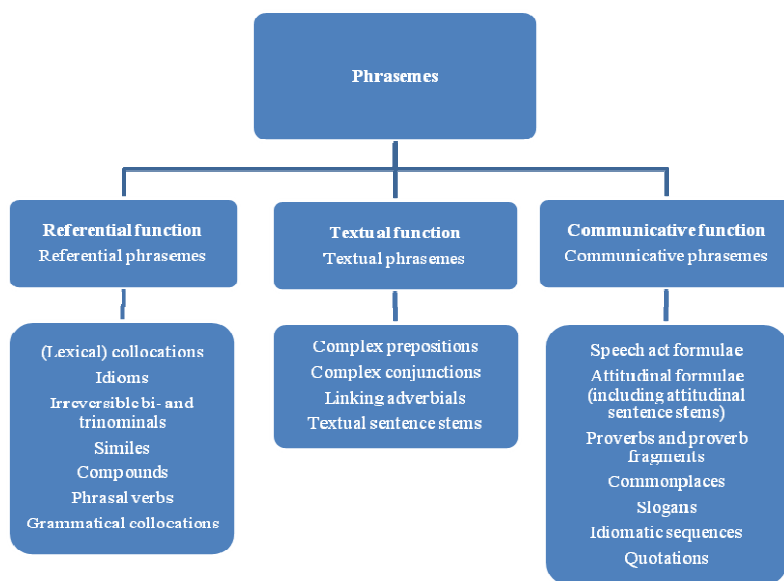


Figure 6: The phraseological spectrum (Granger & Paquot, 2008)

In Granger & Paquot's (2008: 42) proposal, phrasemes are divided into three main categories, i.e. *referential phrasemes*, *textual phrasemes*, and *communicative phrasemes*. Referential phrasemes are "used to convey a content message: they refer to objects, phenomena or real-life facts." The category of textual phrasemes represents an extension of Burger's structural phrasemes (cf. Figure 4) and is used "to structure and organize the content (i.e. referential information) of a text or any type of discourse..." Finally, communicative phrasemes include those expressing "feelings or beliefs towards a propositional content" or explicitly addressing "interlocutors, either to focus their attention, include them as discourse participants or influence them." In their study, Granger & Paquot subsequently give a list of each of the prefabricated units in their respective categories providing each of them with a definition (drawing heavily on the work of major phraseologists, particularly Cowie, Mel'čuk and Burger) and illustration; however, only three elements of the list are relevant for and thus taken into account in the present study, i.e. (lexical) collocations, grammatical collocations, and idioms (for details see Section 4.3.).

4. Collocations

So far, it is clear that collocations represent one type of prefabricated units. Nevertheless, they are also one of the most problematic types of these units for there is no unified approach to this linguistic phenomenon (especially from the point of view of deciding on the status and the relevance of a collocation) and as such there is a number of definitions of collocations varying according to the approach chosen. Furthermore, there are inconsistencies in distinguishing collocations from other prefabricated units. In Section 2.1.2., two approaches to phraseology were described which inevitably leads to two approaches to collocations as well. This section gives a brief description of these two approaches, as well as several other classifications and definitions of collocations (and related concepts) by the major linguists.

4.1. Major approaches to collocations

Traditionally, the term “collocation” has long been used for syntagmatic relations in a language. Nevertheless, since the advent of corpora and corpus linguistics, the analysis of this phenomenon has become more and more prominent yet the state of classification in collocation research is still unclear. It has been understood that collocations are of paramount importance for native speakers and even more important for non-native speakers (especially from the point of view of their acquisition of “collocational knowledge by learning vocabulary not as isolated items but as items in collocation” (Handl, 2008: 46). Handl assumes that the classification of collocations is influenced by various “views on the reason, function, and representation of collocations as habitual lexical co-occurrences in a language” (ibid. 46) and that the problems mainly arise from “the status of collocation as a product of two elements characterised by the varying nature of the relation between the collocational partners” (ibid.). Thus, there are not only different approaches to, or perspectives from, which syntagmatic lexical relations have been approached, but also a wide range of typologies and definitions of collocations as well. The most important of them can be assigned to five major categories which are summarised in the following subsections.

4.1.1. Frequency-based approach

The ‘frequency-based approach’ (Nesselhauf, 2004a), or the ‘statistically oriented approach’ (Herbst 1996: 380), goes back to Firth and has been developed particularly by Halliday and Sinclair. This approach is a text-oriented for it sees collocations as the co-occurrence of words at a certain distance (the syntactic relationship between the elements is not important in order to decide whether they form a collocation or not) and the basic distinction is made between co-occurrences that are frequent, i.e. more frequent than expected if words combined randomly in a language, and those non-frequent. Sinclair (1991: 170) defines collocations as:

the occurrence of two or more words within a short space of each other in a text. The usual measure of proximity is a maximum of four words intervening ... The word whose lexical behaviour is under investigation is called the node [and] it is normally presented with other words to the left and right and these are called collocates. The collocates can be counted and

this measurement is called the “span” [i.e. the measurement, in words, of the co-text of a word selected for study]. A span of -4, +4 means that four word on either side of the node word will be taken to be its relevant verbal environment.

For Sinclair, collocations are primarily lexical co-occurrences of words, though he admits that “this kind of patterning is often associated with grammatical choices as well” (p. 170) and he mentions linguists (e.g. Kjellmer) who include also grammatical relations in their specification of collocation. Next, a distinction is made between “significant collocations,” i.e. co-occurrences of words “such that they occur more often than their respective frequencies and the length of text in which they appear would predict” and “casual,” i.e. non-significant, collocations (Jones & Sinclair, 1974: 21).

In order to clarify the terminology even more, Sinclair (1991: 172) compares collocations with idioms and describes the relationship between them. He defines idioms as “a group of two or more words which are chosen together in order to produce a specific meaning or effect in speech or writing. The individual words which constitute idioms are not reliably meaningful in themselves, because the whole idiom is required to produce the meaning.” As he further explains, “idioms overlap with collocations, because they both involve the selection of two or more words.” Though the line between these concepts is not clear, “we call co-occurrences idioms if we interpret the co-occurrence as giving a single unit of meaning. If we interpret the co-occurrence as the selection of two related words, each of which keeps some meaning of its own, we call it a collocation” (ibid.).

4.1.2. Phraseological approach

The ‘phraseological approach’ (Nesselhauf, 2004a), or the ‘significance oriented approach’ (Herbst 1996: 380), considers collocations as a type of word combination, i.e. an “abstract combination with instantiations in actual texts” (Nesselhauf, 2005: 14), which is “fixed to some degree but not completely” (p. 12). In contrast to the frequency-based approach, the phraseological approach “consistently requires syntactic relations between the elements in order to define word combinations as collocations” (p. 17). As already mentioned above, this approach is strongly influenced by Russian phraseology and the main representatives

of this approach are Cowie, Mel'čuk, and Hausmann. Cowie defines collocations by distinguishing them from other types of word combinations, particularly from free combinations on the one side and idioms on the other. Cowie's attempt to both delimit different types of word combination and define collocations is generally considered one of the most precise (ibid. 14). As shown in Section 3.2. (particularly Figure 1), collocations represent one type of composites (described as having primarily syntactic function) and the distinction between the three types of composites is made on the basis of two main (closely interacting) criteria, i.e. the criterion of transparency, and that of commutability (or substitutability). The criterion of transparency refers to "whether the elements of the combination and the combination itself have a literal or non-literal meaning" (Nesselhauf, 2004a: 14) and the criterion of commutability refers to "whether and to what extent the substitution of the elements of the combination is restricted." Collocations, or more precisely restricted collocations, are thus defined as word combinations in which at least one element has a non-literal meaning, and at least one element is used in its literal sense, and the whole combination is transparent. In addition, some substitution is possible, yet there are arbitrary limitations on substitution. As Nesselhauf (p. 15) points out, there is one inconsistency in Cowie's terminology for the term "collocation" is sometimes used to refer to free combinations as well. In this case, however, Cowie makes a distinction between "open collocations," i.e. free combinations, and "restricted collocations."

4.1.3. Other approaches

In her study, Handl (2008: 49-50) mentions not two but four major approaches to collocations which, however, largely overlap with the frequency-based and phraseological approaches respectively.

The first includes text-oriented definitions of collocations, the second emphasises the associative nature of collocation. Firth (in Palmer, 1968: 181) points out that "it is an order of mutual expectancy..., [i.e. t]here is a certain associative bond between two words that collocate." The next is mainly statistically oriented. The basic question, as Handl puts it, is whether two words "only occur by chance or reappear with greater than random probability" and the last one can be seen as a counter-position to the statistically oriented approaches. Handl calls it the semantic

type for researches having adopted this approach try to “put the relation between co-occurring words down to aspects of meaning.” The two elements of a collocation are called the *basis* and the *collocator* – or *autosemantic* and *syntagmatic components* (Hausmann, 1997: 50) and it leads to a typology of collocations “using semantic features to determine the *collocator*.”

4.2. Criteria commonly used to describe and delimit collocations

As we have seen, there is a large number of definitions (going hand in hand with the various classifications) varying according to the point of view taken towards collocations and on the criteria used (Handl, 2008: 50-53). Basically, there are two main types of criteria, i.e. prerequisites and continua.

The prerequisites are simply conditions that have to be fulfilled in order to treat a word combination as a collocation. The primary and essential criterion for defining collocations is, obviously, the co-occurrence of two or more words (cf. Sinclair). Yet, this criterion brings about a few logical consequences concerning the co-occurring words in question. First, the respective words must be open to combination, i.e. they must belong, for instance, to the “same register or text type (cf. Lipka 2002: 184f), since otherwise they will not usually occur together.” Second, these words must occur in a common context, or more precisely, co-text. Importantly, these words do not have to be in the same clause or sentence in order to be considered a collocation, for it is possible for the two words in question to be separated by other lexical items, cf. Greenbaum’s (1970: 11) example with the collocation *collect stamps*:

- (1) a. *They collect many things, but chiefly stamps.*
b. *They collect many things, but [they] chiefly [collect] stamps.*

As Handl further explains, the only condition that must be fulfilled is that the “syntactical relation between the constituents in question allows a reconstruction of an adjacent collocation” (ibid.) as given in (1) b. In contrast, example (2) shows the same elements as in (1), but here these two elements do not form a collocation (the collocation is in this case *collect revenue*), cf. Greenbaum (ibid.):

- (2) *The first adhesive postage stamp was used in Great Britain in 1840. At the time, the British post office was having trouble collecting revenue.*

The second criterion according to Handl (p. 51), i.e. continua, is more complicated for it represents a gradable phenomenon, i.e. it is not a question whether a continuum can be applied or cannot, but rather they can be applied to varying degrees to different kinds of collocations.

Semantic transparency (closely connected with the notion of compositionality) is the first continuum and its importance lies in the fact that it is responsible for the distinction between collocations and idioms and as such it can be seen as the counterpart to idiomacity (ibid. 51). Nevertheless, the boundary between collocations and idioms is rather fuzzy for it depends on the degree of semantic transparency of a word combination in question. Thus, in other words, there are different degrees of transparency (or opacity) depending on the “semantic contribution an element makes to the meaning of the whole expression” (Handl, p. 52).

The second continuum is the so-called collocational range, i.e. the number of potential collocates a node can take. The collocational range can be either restricted or wide; however, the wider the list of potential combinations, the less typical the collocation (i.e. the widest ranges usually suggest a word combination being a free combination), and vice versa, i.e. the more restricted range, the higher probability of the whole combination being either idiom or a complex lexeme, cf.:

- (3) a. *to face* + *the facts / truth / problems / reality, etc.*
- b. *to face* + *charges / counts*
- c. *to face* + *the music*

Example (3) clearly shows a narrowing of the collocational range, i.e. (3a) can be undoubtedly considered a collocation because of its collocational range, whereas (3c) inevitably represents an idiom for there is only one possible combinatory element. (3b) is, according to Handl, a sort of “transition area.”

The third continuum for classification of collocations is the frequency of co-occurrence, and as already mentioned above, it is essential not only from the point of view of corpus linguistics, but also the frequency of co-occurrence is of a prime importance in order to decide on the relevance of a collocation from the point of view of learners of a language. Nevertheless, the frequency of co-occurrence alone is not a reliable criterion. As Handl points out, further statistical aspects taking into account questions of probability and interrelation between the elements also have to

be incorporated. These statistical aspects together with the collocational range and the frequency of co-occurrence can be used to derive a fundamental criterion for collocations, i.e. the predictability or mutual expectancy of words. Predictability, or mutual expectancy, can be described as a psychological or cognitive feature which is decisive for collocations, especially from the point of view of native speakers who often become aware of collocations when they are used inappropriately or creatively in spoken or written language.

4.3. The definition and classification of collocations in this study

The approach to collocations in the present study is based upon Granger & Paquot's (2008) paper, particularly for two main reasons. First, they try to reconcile various approaches to (and their respective definitions of) collocations into one spectrum (for details see Section 3.3.), and second, the present study is built upon Granger's (1998) research study of collocations. Thus, the definition and classification of collocations in this study has been directly adopted from the phraseological spectrum proposed for the linguistic classification of prefabricated units (Granger & Paquot, p. 43). The category of idioms is presented for comparison and distinguishing them from collocations:

Category	Definition and illustration
(Lexical) collocations	(Lexical) collocations are usage determined or preferred syntagmatic relations between two lexemes in a specific syntactic pattern. Both lexemes make an isolable semantic contribution to the word combination but they do not have the same status. Semantically autonomous, the "base" of a collocation is selected first by a language user for its independent meaning. The second element, i.e. the "collocate" or "collocator," is selected by and semantically dependent on the "base." Examples: <i>heavy rain, closely linked, apologize profusely.</i>
Grammatical collocations	Grammatical collocations are restricted combinations of a lexical and a grammatical word, typically verb / noun / adjective + preposition, e.g. <i>depend on, cope with, a contribution to, afraid of, angry at, interested in.</i> The term "grammatical collocation" is borrowed from Benson et al. (1986) but our definition is slightly more restricted as these authors also use the term to refer to other valency patterns, e.g. <i>avoid + -ing</i> form, which we do not consider to be part of phraseological spectrum.
Idioms	The category of idioms is restricted to phrasemes that are constructed around a verbal nucleus. Idioms are characterized by their semantic non-

compositionality, which can be the result of a metaphorical process. Lack of flexibility and marked syntax are further indications of their idiomatic status. Examples: *to spill the beans*, *to let the cat out of the bag*, *to bark up the wrong tree*.

Table 1: Categories of referential phrasemes (extract) (Granger & Paquot, 2008: 43)

In the present study, only lexical collocations are analysed, in particular, those consisting of two elements: an amplifier plus an adjective, i.e. amplifiers ending in *-ly* and functioning as pre-modifiers to their respective adjective. These combinations fully reflect those analysed in Granger's study (for details see Sections 5.1. and 6. – Material and Method).

4.3.1. Amplifiers

Amplifiers (Quirk et al., 1985: 590) represent one category of intensifying adjectives or adverbs respectively (though some prepositional phrases or noun phrases can also be used as amplifiers). Adjectives used as amplifiers have a heightening effect on the noun they modify, i.e. they “scale upwards from an assumed norm” (ibid. 429) and are either central or attributive only, i.e. *a complete victory / The victory was complete*, or *great destruction / The destruction was great* in contrast to *a complete fool / *The fool is complete* (ibid, 429). Adverbs used as amplifiers may either modify adjectives or verbs. They, again, “scale upwards from an assumed norm” (ibid. 445), cf. *funny film* vs. *a very funny film*. Importantly, these adverbs (functioning as amplifiers) can co-occur only with items which are gradable, i.e. referring “to a quality that is thought of as having values on a scale” (ibid. 469). Furthermore, it is possible to contrast most amplifiers in “alternative negative with *to some extent*, and this propensity is a semantic test for their inclusion in the class of amplifiers” (ibid. 590). Examples (4a) and (5a) for instance are in such contrast compared to the category of emphasizees which cannot be used in this way (Examples (4b) and (5b)), cf.:

- (4a) He didn't ignore my request *completely*, but he did ignore it *to some extent*.
- (4b) *He didn't *really* ignore my request, but he did ignore it *to some extent*.
- (5a) They don't admire his music *greatly*, but they do admire it *to some extent*.
- (5b) *They don't *definitely* admire his music, but they do admire it *to some extent*.

There are two sub-classes of amplifiers, i.e. maximizers and boosters. Maximizers can denote the upper extreme of the scale, whereas boosters denote a high degree, or a high point on the scale (ibid. 590). Common maximizers are for instance *absolutely, altogether, completely, entirely, extremely, fully, perfectly, thoroughly, totally, utterly*, etc. Common boosters include for instance *badly, bitterly, deeply, enormously, greatly, highly, intensely, much, severely, strongly*, etc. As Quirk et al. (p. 591) point out, the distinction between these two subclasses is not fast; particularly, when “maximizers are in *M* position they often express a very high degree, whereas when they are in *E* position they are more likely to convey their absolute meaning of extreme degree.”

5. Previous research on collocations in learner English

With the recognising of the importance of collocations, a number of studies analysing this phenomenon have been published. Generally speaking, these studies are based on either elicitation tests (i.e. cloze tests or translation tasks) or production data, and focus on the production of collocations rather than on the comprehension of them since the production of collocations is much more problematic for learners than the comprehension. Nevertheless, there are two major problems with most of the studies. First, in practice, collocations are very often not carefully delimited from other types of word combinations so that compounds or even idioms are sometimes included in the analysis without further discussion. Second, though there is the shared conclusion of these studies that the production of collocation presents a problem for second language learners, only a few studies provide further analysis of the results (Nesselhauf, 2005: 3, 4, 7).

The results of studies on collocation vary according to the method of investigation chosen yet some results have emerged. As Nesselhauf (2005: 3-8) points out, majority of these results mainly confirm the observation that collocations indeed present a problem for second language learners. Learners are insecure in the use of collocations and, interestingly, according to several studies there is no correlation between the general proficiency of a learner and the number and the acceptability of the collocations used. In addition, it has been proved that mere exposure to

collocations does not usually lead to their acquisition and that reading only slightly improves learners' knowledge of collocations. Most importantly, studies have also shown that there is a strong L1 influence; it has been shown that the translation of collocations from L2 to L1 is usually unproblematic for learners whereas the translation from L1 to L2 proved to be considerably more difficult for learners (and the translation of verbs as parts of collocation proved to be more problematic than translating other lexical items). Similarly, non-native-like collocations are based on transfer from L1 to L2 in approximately a tenth of cases. Furthermore, learners in general used fewer collocations than native speakers (e.g. in several studies large number of accurate, or acceptable word combinations were not marked by the learners), but that they greatly overuse a small number of them (in particular those that are frequent in English and / or similar to an L1 combination). Nesselhauf (2005) comes to similar conclusions, i.e. she finds that whereas learners in general use smaller number of prefabricated units than native speakers, they did use quite a large number of native-like collocations, though it "cannot be assumed that all of them were stored and produced as chunks" (p. 247). Also, in the use of chunks, learners vary much more than native speakers and "frequently appear to create collocations using individual bricks" for the links between chunks and meaning are weaker in learners' mental lexicons as compared to native speakers (p. 247, 248).

To sum it up, in both types of studies collocations have been proved to be a serious problem for learners of English. The studies have generally shown that learners use fewer collocations than native speakers, except for a small number of frequent collocations which are overused. Furthermore, learners are insecure in the production of collocations (which is closely connected to the fact that collocation problems are more serious than general vocabulary problems) and they are often not aware of restrictions yet at the same time they are also not aware of the "full combinatory potential of words they know (Nesselhauf, 2005: 8). Since the present study is based on Granger's (1998) study on collocations, in the following subsection, relevant parts and findings of this study are described in more detail.

5.1. Granger' Contrastive Interlanguage Analysis

As already mentioned above, Granger's (1998) research is of a crucial importance from the point of view of present paper for the research part is based upon it. The main aim of Granger's study is to "uncover factors of non-nativeness in advanced learners' writing" (p. 146), particularly to analyse and compare the use of (or the ability to use) collocations and formulae by native and non-native speakers of English. Granger begins her article by summarising the main reasons for the growing interest in the use of prefabricated units in EFL, i.e. the promotion of the syntagmatic investigation of lexis due to the emergence of lexico-grammar inspired by Halliday and Sinclair; the development of corpus linguistics providing linguists with computational means to analyse and uncover various lexical patterns; and finally, the establishment of pragmatics as a major field of study in its own right in linguistics as well as in EFL (p. 144).

The methodology employed for Granger's research is (what she herself calls) the Contrastive Interlanguage Analysis (CIA). There are basically two methods or types of comparison used in CIA: a comparison between native and non-native varieties of the same language (i.e. L1 vs. L2), or a comparison of several non-native varieties (i.e. L2 vs. L2). In her 1998 study, the former type of analysis is employed. The data Granger compares come from two corpora, native and learner corpus respectively. The native speaker corpus (NS) contains three main parts, i.e. the Louvain essay corpus, the student essay component of the International Corpus of English (ICE), and the Belles letters category of the Lancaster-Oslo-Bergen corpus (LOB). The learner (i.e. non-native, NNS) corpus is a corpus of writing by advanced French-speaking learners of English, particularly a subcorpus of the ICLE database. The initial hypothesis was that learners would use prefabricated (or conventionalised) language in a much lesser extent than native speakers since the use of such language is "universally presented as typically native-like." She expected learners to make use of individual bricks (cf. Nesselhauf 2005, or Kjellmer 1991) rather than prefabricated sections.

A collocation in Granger's study consists of two elements and it is defined as "the linguistic phenomenon whereby a given vocabulary item prefers the company if another item rather than its 'synonyms' because of constraints which are not on the level of syntax or conceptual meaning but on that of usage" (p. 146). These

combinations are called either “lexical collocations” or “restricted collocations.” In particular, for the collocation studied, Granger selected one category of intensifying adverbs, i.e. amplifiers ending in *-ly* and functioning as modifiers, such as *perfectly natural*, *closely linked*, or *deeply in love* etc. As Granger (p. 147) points out, these word-combinations constitute “a particularly rich category of collocation, involving as they do a complex interplay of semantic, lexical, and stylistic restrictions and covering the whole collocational spectrum, ranging from restricted collocability – as in *bitterly cold* – to more open collocability – as in *completely different / new / free* etc.” In the excerpt, all words ending in *-ly* were automatically retrieved via the text-retrieval software TACT from both corpora and subsequently sorted according to Granger’s predefined semantic and syntactic criteria.

The mere comparison of the number of types and tokens in the two corpora revealed the first important finding, i.e. the statistically significant underuse of amplifiers in the NSS corpus (both in the numbers of types and tokens). The next step was to find out whether this underuse was general or related only to particular amplifiers or categories of amplifiers. The data revealed that only three individual amplifiers demonstrated statistically significant results, i.e. *completely* and *totally* were overused by learners in comparison to native speakers, and *highly* was underused in the learners’ writings. In addition, since these amplifiers were found in combination with a wide variety of words, these amplifiers can be seen as “all-round amplifiers or ‘safe-bets’” (p. 148) for practically all these combinations were felt to be acceptable by native speakers. Granger suggests that the reason of the overuse or underuse of these amplifiers can be found in the lexical correspondences between the two languages studied, i.e. the amplifiers *completely* and *totally* respectively have direct translation equivalent in French (i.e. *complètement* and *totalelement*) and these French equivalents display “similarly few collocational restrictions.” In contrast, the French equivalent for *highly*, i.e. *hautement*, is only used in formal French and it is not very frequent which can be seen as a sufficient explanation of the underuse of this amplifier in the learners’ writings.

For the examination of amplifiers according to categories, Granger divided amplifiers into “maximizers” (expressing the highest degree) and “boosters” (expressing merely a high degree) (cf. Quirk et al., Section 4.3.1.). The data revealed that in the category of maximizers, the number of types in the learner corpus is the same as in the native one, and the number of tokens in the former is slightly higher

(mainly due to overuse of *completely* and *totally*) than in the latter. Yet, as Granger admits, the overall figures are not statistically significant. However, in the category of boosters, the data revealed an underuse of boosters in the learners' writings which is "significant enough to explain the general underuse of amplifiers attested to earlier" (p. 148).

The category of boosters represents 66 per cent of the amplifiers in the NS corpus whereas this category represents only 35 per cent of all amplifiers found in the NNS corpus. The number of types in this category is also much higher than in the category of maximizers (because of the category of boosters represents an opened set (p. 149). Yet, by subdividing the category of boosters into three categories, i.e. boosters that are exclusively used by the native speakers, those that are exclusively used by the non-native speakers, and those that are common to both groups, Granger revealed further differences in the use of boosters by native speakers and learners. Interestingly, the majority of the boosters (77.5 per cent) used by non-native speakers were used by native speakers as well, whereas the majority of the boosters (63 per cent) found in native writings were used exclusively by natives. The "native-exclusive" combinations were either stereotyped combinations (e.g. *actually aware*, *readily available*, *painfully clear*, *vitaly important*, etc.) or creative combinations (such as *ludicrously ineffective*, *monotonously uneventful*, etc.), and both of these types were significantly underused by the learners. Interestingly, the few stereotyped combinations found in the NNS corpus either have, again, direct translation equivalents in French or these combinations are typical ones in both English and French.

Thus, the first part of Granger's study confirms the initial hypothesis concerning learners' lower use of prefabricated units than native speakers. Importantly, it has been shown that most collocations used by the learners have direct translation equivalents in L1 and thus may be the result of the transfer from L1 to L2. However, as Granger points out, the learners indeed seem to use amplifiers "more as building bricks rather than as parts of prefabricated sections" (p. 151). Crucially, some amplifiers are used as "general-purpose" items which is further supported by the independent analysis of *very*, the "all-round amplifier *par excellence*" (ibid.). This analysis revealed a highly significant overuse of this amplifier and it can be postulated that the underuse of *-ly* amplifiers in learners' writings is compensated for by the overuse of *very* by the learners.

In the second part of her study, Granger focuses on the notion of a significant collocation. It has been shown that learners do use collocations but their underuse of native-like collocations as well as atypical word-combinations in L2 can be connected with an “underdeveloped sense of salience and of what constitutes a significant collocation” (p. 152). In this study, Granger extracts introspective data on collocations by submitting a word combination test to 56 French learners of English and 56 native speakers of English. The subjects were asked to choose the acceptable collocates of particular amplifiers.³ Consequently, Granger confined her attention to the collocates which were more frequently associated with the amplifier than all the others from the same group (marked with asterisks in the study by both native- and non-native speakers of English). These combinations were considered particularly salient in the subjects’ minds and revealed some interesting results. Overall, learners marked with an asterisk 280 combinations as compared to 384 combinations marked by native speakers, suggesting the learners’ weak sense of salience. For instance, *readily available* was selected by 43 native speakers in contrast with only 8 learners. Similarly, *bitterly cold* was asterisked by 40 native speakers but only by 7 non-native ones. In case of *blissfully*, the native speakers chose either *blissfully happy* or *blissfully ignorant*, whereas there were only four cases of the former and even no case of the latter combination found in the non-native responses. Furthermore, the fact that the learners marked a higher number of possible collocations indicates that the learners not only have a weak sense of salience, but also that their sense of salience is rather misguided. Even though in some combinations the learners showed a good sense of salience, e.g. in case of *fully aware*, *fully reliable* or *highly significant*, there was a number of other combinations which were selected as significant by the learners yet none of which was considered significant by the native speakers, such as *fully different*, *fully significant*, *highly impossible*, or *highly difficult*, etc.

According to Granger, introspective data not only can play a role in revealing features of learner language, but also they can “be valuable in providing a

³ The amplifiers presented were: highly, seriously, readily, blissfully, vitally, fully, perfectly, heavily, bitterly, absolutely, utterly and their possible collocates were as follows: significant / reliable / ill / different / essential aware / miserable / available / clear / happy / difficult / ignorant / impossible / cold / important.

clear notion of what constitutes a significant collocation” (p. 154). Her study has demonstrated that non-native speakers possess “severely limited phraseological skills” in the sense that they use “too few native-like prefabs and too many foreignersounding ones” (p. 158). From the point of view of the present study of this study, it is supposed that its results would be similar to those of Granger’s and it is hoped that the present research will reveal some specifics of Czech learners of English relating to the use and knowledge of collocations.

6. Material and Method

In addition to Granger's (1998) study, the main secondary texts used were the main English grammars, e.g. *A Comprehensive Grammar of the English Language* (1985) and *The Cambridge Grammar of the English Language* (2002). As a complementary source, Libuše Dušková's *Mluvnice současné angličtiny na pozadí češtiny* (2006) was used.

The present study is based on the results of two types of tests given to the Czech learners of English focusing on their knowledge of and ability to use English collocations compared to native speakers' data extracted from the *British National Corpus*. The non-native speakers' data for the analysis were collected from Cloze tests and Significance of collocation tests. The samples of both tests are presented in the Appendix (see p. 90-98).

6.1. The Cloze Test

The Cloze test was based on the elicitation of acceptable collocations consisting of an amplifier and an adjective. There were ten participants, i.e. ten non-native speakers, of the Cloze test. There were two males and eight females and the average age of the participants was twenty-six. They have been learning English for sixteen years on average and the average length of time spent in an English speaking country was nine months. The instruction was to complete each of the 20 sentences with at least one and a maximum of 10 amplifiers, i.e. adverbs ending in *-ly* expressing a high degree (e.g. *The news was -----ly good for me*). The twenty adjectives chosen for the test were selected from Leech, Rayson & Wilson's (2001) *Word Frequencies in Written and Spoken English: based on the British National Corpus*, in particular from its Companion Website.⁴ The adjectives were manually selected from the Frequency lists, namely "List 5.3: Frequency list of adjectives (by lemma)."⁵ In order to avoid a bias towards particular amplifiers, an attempt to

⁴ Leech, Rayson & Wilson. *Companion Website for Word frequencies in Written and Spoken English* <<http://ucrel.lancs.ac.uk/bncfreq/>> (accessed 17 June 2012)

⁵ Leech, Rayson & Wilson. *Companion Website for Word frequencies in Written and Spoken English*, "List 5.3.: Frequency list of adjectives (by lemma)" <http://ucrel.lancs.ac.uk/bncfreq/lists/5_3_all_rank_adjective.txt> (accessed 17 June 2012)

include only the most neutral sentences in the test was made using the *Sketch Engine*, a Corpus Query System incorporating word sketches, one-page, automatic, corpus-derived summary of a word's grammatical and collocational behaviour.⁶ For each adjective, its most frequent collocations were found via Word Sketch. Acceptable amplifiers for each adjective were shown in the "Modifier" section and those with the highest frequencies were selected for further processing. The next step was to find the most neutral, i.e. the most frequent, co-texts or words co-occurring with the most frequent collocations (i.e. combinations of an amplifier plus an adjective). This was made via the option "Konkordance → Fráze (i.e. the collocation in question) → Vytvořit konkordanci." Then, the actual co-texts of the whole combination were displayed and subsequently, words co-occurring with the combinations were sorted by using the option "Kokace (atribut: word, rozsah: <-5, +5>, minimální frekvence v korpusu: 5, minimální frekvence v daném rozsahu: 3) → Vytvořit seznam → Frekvence." From these data, the most neutral sentences were compiled (at least the core of the sentences) which were then cross-checked with the *BNC* in order to confirm the high occurrence of those word combinations. When completed, the Cloze test was given to the participants and the data obtained from the tests were sorted according to the frequencies of amplifiers suggested by the non-native speakers and proceeded to further analysis and comparison with the *BNC* data.

Since there were no native speakers participating in these two tests, it was necessary to gain native speakers' data from the *British National Corpus*. The excerption was performed by a KWIC search. For the Cloze test, particular adjectives used in the sentences were used as a key phrase, the "Collocation" box was set to <-1, 0> and in the "POS List" box, the option "adv.all" was selected. From the resulting list, the five most frequent amplifiers were selected manually and sorted according their actual frequencies in the corpora. For one part of the analysis, it was necessary to find out whether the combinations suggested by the non-native speakers are acceptable in English, i.e. whether these combinations actually exist in English or not. In these cases, the excerption of the data was performed by a simple text search, using the particular combinations as a key phrase.

⁶ *Sketch Engine* <<http://www.sketchengine.co.uk/>> (accessed 18 June 2012)

6.2. The Significance of Collocation Test

The Significance of collocation test was directly adopted from the Granger's (1998) study. Unlike in the Cloze test where suitable amplifiers had to be supplied, here the participants of the test were presented with 11 amplifiers each with its own set of 15 adjectives (e.g. **highly**: *significant reliable ill different essential aware miserable available clear happy difficult ignorant impossible cold important*) and were asked to choose the acceptable collocates of the amplifiers by underlining all the adjectives which in their opinion go well with the amplifier. Moreover, they were requested to circle the adjective which in their opinion was associated with the amplifier more frequently than all the others (i.e. one particularly salient adjective for each amplifier). The data obtained from the tests was sorted according to the frequencies of adjectives suggested by the non-native speakers and proceeded to further analysis and comparison with the data extracted from the *BNC* (see below).

There were fifty participants, i.e. fifty non-native speakers, in the Significance of collocation test. There were sixteen males and thirty-four females and the average age of the participants was twenty-one. They have been learning English for twelve years on average and the average length of time spent in an English speaking country was three months (in fact, only eight participants have spent one year or more in an English speaking country, thus it can be said that the participants of this test have had very little or no exposure to the L2 in general).

In order to obtain native speakers' data for comparison with the results of the Significance of collocation tests, the same search method was used. For this test, it was the amplifiers which were used as key phrases. The "Collocation" box was set to <0, +1> and in the "POS List" box, the option "adj.all" was selected. From the resulting list, the frequencies of the combinations of an amplifier and the fifteen adjectives presented as options in the test were chosen and recorded for further procession. The data collected for each test was arranged into the native speaker database and used for comparison with the non-native speakers' data.

Furthermore, it was also necessary to extract the most frequent Czech amplifiers in order to explain some specifics of Czech learners of English relating to the use of particular collocations. These amplifiers were extracted via Český národní korpus (the *Czech National Corpus*) which is an academic project focusing on

building a large electronic corpora of both written and spoken Czech.⁷ The excerption was made by a simple text search, the command used was: [tag="D.*"][tag="A.*"]. The most frequent amplifiers were then selected manually via the frequency distribution.

⁷ Český národní korpus (FFUK). "What is a Corpus." <www.korpus.cz/english/co_je_korpus.php> (accessed 23 June 2012)

Part II – Data Analysis

7. Analysis – Introduction

This part of the study is based on the results of Cloze tests and Significance of collocation tests. It is divided into four main sections, including a section which describes English amplifiers in general, one which contains pedagogical implications concerning the prefabricated nature of languages and Section 8. which summarises the results presented in this part and points out the main findings resulting from the analysis.

7.1. English amplifiers: General overview

The initial hypothesis for this study was that learners' knowledge of and ability to use collocations would be much weaker in comparison with native speaker data. In particular, it was presupposed that learners would make less use of collocations in the investigated data (and thus of course in their actual use of English, both spoken and written) than native speakers, that the preferred amplifiers would differ in comparison with native speaker data, and that learners would make a number of mistakes in their use of collocations. The results of both tests, i.e. the Cloze test as well as the Significance of collocation test, are presented and analysed in the following sections.

As a first step allowing evaluation of the non-native speakers' use of amplifiers, all possible combinations of an amplifier plus an adjective were extracted from the *BNC*. Table 2 shows the first fifteen most frequent English amplifiers followed by an adjective in general:

No.	AMP + adj.		BNC	
			Σ	%
1.	<i>particularly</i>	+ adj.	8370	8.86
2.	<i>really</i>	+ adj.	8175	8.66
3.	<i>highly</i>	+ adj.	6545	6.93
4.	<i>extremely</i>	+ adj.	5948	6.30
5.	<i>fairly</i>	+ adj.	4661	4.93
6.	<i>totally</i>	+ adj.	3318	3.51
7.	<i>completely</i>	+ adj.	3303	3.50
8.	<i>absolutely</i>	+ adj.	2994	3.17
9.	<i>entirely</i>	+ adj.	2851	3.02
10.	<i>perfectly</i>	+ adj.	2530	2.68
11.	<i>especially</i>	+ adj.	2386	2.53
12.	<i>fully</i>	+ adj.	2070	2.19
13.	<i>apparently</i>	+ adj.	1623	1.72
14.	<i>clearly</i>	+ adj.	1558	1.65
15.	<i>truly</i>	+ adj.	1518	1.61
TOTAL			94448	61.25

Table 2: Raw frequencies of amplifiers based on the *BNC*

As we can see in Table 2, the first fifteen most frequent amplifiers followed by an adjective comprise more than sixty per cent of all possible combinations of an amplifier plus an adjective found in the *BNC*. The most frequent English amplifiers are *particularly*, *really*, *highly*, and *extremely* comprising almost one third of all existing combinations. All these amplifiers display open collocability, i.e. ranging from 676 existing combinations of *truly* + an adjective to 1438 existing combinations *particularly* + an adjective. Thus, it can be claimed that these amplifiers are used as – what Granger calls – all-round amplifiers or “safe bets.”

7.2. Results of the Cloze test

7.2.1. Overall results

According to the results of the cloze tests and the excerpt from the *BNC*, the overall number of amplifiers analysed in this part of the study was forty-eight. The resulting number combines amplifiers suggested by the non-native speakers in

cloze tests and five most frequent amplifiers for the respective adjectives extracted from the British National Corpus. Since the amount of data supplied by the non-native speakers and the respective sample of the data extracted from the *BNC* was incomparable, the comparison of the non-native and native speakers' data is expressed as a per centage of the total where necessary.

Thus, in this part, the first step was to make a list of all suggested combinations found in the cloze tests and these combinations were subsequently supplemented with the five most frequent collocations of the respective adjectives subjected to the analysis. Then, the number of existing / suggested combinations of all forty-eight amplifiers as well as their actual occurrences in both the cloze tests and the *BNC* were counted. In total, there were twenty adjectives which were preceded by forty eight different amplifiers, resulting in one hundred and eighty-eight different combinations:

Overall number of adjectives	20
Overall number of amplifiers	48
Overall number of combinations (an amplifier + an adjective)	188
Overall number of occurrences of these combinations	9914

Table 3: Overall results of the excerption

The next step was to compare the number of amplifiers, existing / suggested combinations of these amplifiers plus the respective adjective, and the actual number of occurrences for both databases, i.e. the native speaker (NS) database and the non-native speaker (NNS) one. Though the amount of data is markedly different in both databases, the comparison revealed some general substantial differences in the use of collocations by native- and non-native speakers of English:

Amplifiers	NS		NNS	
	Σ	%	Σ	%
Amplifiers	37	77	33	69
Combinations	102	54	149	80
Occurrences	9612	97	302	3

Table 4: Raw frequencies of amplifiers based on the NS and NNS data

Though it was expected that the number of amplifiers found in the NS database would be markedly higher than that of the NNS database, the results of the comparison showed that the number of amplifiers in both databases was rather similar. The NS database contained thirty-seven amplifiers, comprising almost eighty per cent of all possible amplifiers (77 %) whereas the non-native speakers made use of thirty-three amplifiers, i.e. almost seventy per cent of all amplifiers from the database. Thus, the underuse of amplifiers in the NNS category was not as striking as expected. Nevertheless, one must take into account that the NS database contained only five most frequent amplifiers for each adjective, i.e. if we compared the NNS results with all possible amplifiers for each adjective (not only five most frequent ones), there would be a significant underuse of amplifiers in the NNS database simply due to the amount of data compared. Since only five most frequent amplifiers for each adjective were extracted from the *BNC*, these amplifiers can be described as those used rather generally, they would often be repeated, and there would be presumably a lower number of stereotyped combinations as well as “creative” combinations with these amplifiers.

Yet, even though there was no significant underuse of the NNS amplifiers in the present data, these two figures revealed the first important thing. Considered that there were forty-eight amplifiers in total and the NS database contained thirty-seven of them, there were eleven amplifiers used by the non-native speakers which were not found among five most frequent amplifiers of the NS database. Similarly, since the NNS database contained thirty-three amplifiers, there were fifteen amplifiers which were found among the most frequent ones in English yet not suggested by the non-native speakers. In other words, these figures suggest that in general there were striking differences between the use or selection of amplifiers in the NS and the NNS database (for details see below).

Moreover, the overall results show another important thing, i.e. the statistically significant difference between the existing native combinations and combinations suggested by the non-native speakers. Whereas there were one hundred and two (most frequent) combinations in the NS database, the NNS database contained one-hundred and forty-nine combinations, i.e. eighty per cent of the total (though the number of suggested amplifiers is slightly lower). Thus, if we compare the number of amplifiers and their combinations in both NS and NNS data, we found that whereas in the NS data each amplifier was found in combination with almost

three different adjectives on average, in the NNS data each amplifier was found in combination with more than four different adjectives. In other words, it suggests that in contrast to native speakers, the non-native speakers generally used the same amplifiers regardless of the following adjective, revealing not only that the sense of salience was somewhat weaker or underdeveloped in the case of the non-native speakers but also that the “creative” ability of the non-native speakers was significantly limited, i.e. they were in fact not able to come up with different amplifiers for particular adjectives and the suggested sets of amplifiers were very often repeated as if the following adjectives did not play role (especially in the case of amplifiers or adjectives displaying open collocability).

On the basis of the results of raw frequencies of amplifiers extracted from the NS and NNS database, the next step was to establish whether the underuse of amplifiers by the non-native speakers was a general phenomenon or due to underuse of particular amplifiers. On the whole, the frequencies of the individual amplifiers were often too low, i.e. not significant enough, for meaningful conclusions to be drawn. However, as shown in Table 5, five of the amplifiers demonstrated statistically significant differences:

Amplifier	NS occurrences		NNS occurrences	
	Σ	%	Σ	%
<i>absolutely</i>	198	2.1	32	10.6+**
<i>truly</i>	61	0.6	17	5.6+*
<i>incredibly</i>	0	0	14	4.6+*
<i>utterly</i>	10	0.1	13	4.3+*
<i>particularly</i>	1780	18.5	10	3.3-**
TOTAL	9612	100	302	100

Table 5: Individual amplifiers demonstrating statistically significant differences in occurrences

Note: Significant levels of overuse or underuse on the part of learners are indicated by a plus or a minus sign followed by an asterisk (in case of very significant levels of overuse or underuse there is a double asterisk).

As can be seen, *absolutely* demonstrated very significant level of overuse on the part of the non-native speakers. One of the reasons for the overuse of *absolutely* may well be that this amplifier has a direct translation equivalent which is frequent in Czech – *absolutně* – and which displays similarly few collocational restrictions. Even though other amplifiers demonstrating statistically significant overuse, i.e. *truly*, *incredibly* and *utterly*, do not have direct translation equivalents in Czech, they translate very nicely into Czech and these translation equivalents are also frequent ones which, again, can be seen as one of the reasons for their overuse in the NS database.

In contrast, out of the forty-three remaining amplifiers, i.e. those with statistically low frequencies, more than sixty per cent of amplifiers were underused by the non-native speakers in contrast to native speakers' data, hence the little underuse of amplifiers on the part of the non-native speakers in general. *Particularly* was the only amplifier demonstrating a very significant level of underuse on the part of the non-native speakers. One possible explanation for its underuse may be that this amplifier has actually two meanings, i.e. *especially*, or *more than usual*,⁸ and it is the first meaning of this amplifier which is more associated with it in Czech, i.e. *zejména*, *hlavně*, *především* (rather than *obzvlášť*, or *mimořádně*), and as such does not fit into most of the cloze test sentences when translated into Czech.

Subsequently, it was crucial to find out whether the higher number of suggested combinations in the NNS database in comparison with the combinations in the NS database was a general phenomenon or due to particular amplifiers. Overall, there were twenty amplifiers in the database which were found in combination with more adjectives in the NS database than the NNS database. In other words, there were more existing combinations than the non-native speakers suggested. Yet, on average, there were only 1.6 more combinations for each amplifier in the NS database than in the NNS database, i.e. in the case of twenty amplifiers, the non-native speakers suggested usually one or two combinations less than possible. In contrast, there were twenty-four amplifiers in the database which were found in combination with more adjectives in the NNS database. However, on average there were 3.3 more combinations for each amplifier in the NNS database in comparison to the NS database, hence the higher number of types on the part of non-native speakers. Only four amplifiers had the same number of types in the NS and the NNS

⁸ *Cambridge Advanced Learner's Dictionary* (2006), 2nd edition (electronic version).

database. Interestingly, Table 6 shows the amplifiers demonstrating significant differences between the number of combinations in the NS and NNS database (interestingly, the table contains all of the amplifiers presented in Table 5, i.e. those demonstrating the significant levels of overuse or underuse respectively – thus these two tendencies are clearly interconnected):

Amplifier	NS combinations	NNS combinations
<i>absolutely</i>	4	9+*
<i>completely</i>	4	8+*
<i>fairly</i>	0	6+**
<i>highly</i>	2	7+*
<i>incredibly</i>	0	10+**
<i>particularly</i>	13	7-**
<i>simply</i>	0	4+*
<i>truly</i>	2	11+**
<i>unbelievably</i>	0	4+*
<i>utterly</i>	1	10+**
TOTAL	102	149

Table 6: Individual amplifiers demonstrating statistically significant differences in the number of NS and NNS combinations

Note: Significant levels of overuse or underuse on the part of learners are indicated by a plus or a minus sign followed by an asterisk (in case of very significant levels of overuse or underuse there is a double asterisk).

As we can see, the difference between the number of combinations in the NS and NNS databases are, at least in some cases, indeed striking. These figures support the already mentioned claim that the non-native speakers use fewer amplifiers than the native speakers; however, they use them rather generally, i.e. regardless of the following adjective / collocation. In other words, the majority of amplifiers used by the non-native speakers were treated as amplifiers demonstrating open collocability

(which is not generally true), their sense of salience regarding particular collocations was significantly weaker and also their “creative” ability was underdeveloped. Again, in the case of all of the amplifiers found with the highest number of combinations suggested by the non-native speakers, there was an obvious and strong L1 influence. Thus, so far it can be claimed that the non-native speakers seem to use amplifiers more as building bricks than as parts of prefabricated units for there is an evident influence of their mother tongue, i.e. Czech, on the selection of particular amplifiers, i.e. the amplifiers with direct translation equivalents or those which translate nicely into Czech were the most frequent ones in the NNS database, as if there were no differences between English and Czech, i.e. as if there was a direct, straight way (one-to-one relationship) of translating these combinations from English to Czech or vice versa.

7.2.2. Non-native speakers: (non)attaining the native “ideal”

So far, the general differences between the native and non-native use of collocation have been shown and described. This section is of a vital importance for its main aim is to find out to what extent – or if at all – the non-native speakers attain the native “ideal,” i.e. how the selections as well as orders of particular combinations suggested by them correspond with the existing native ones.

The comparison of the most frequent amplifiers used by native speakers to those suggested by non-native speakers also revealed first important differences. Tables 7 and 8 show the first ten most frequent amplifiers in both databases:

Amplifier	NS occurrences	
	Σ	%
1. <i>really</i>	1934	20.1
2. <i>particularly</i>	1780	18.5
3. <i>extremely</i>	1287	13.4
4. <i>completely</i>	549	5.7
5. <i>totally</i>	418	4.3
6. <i>especially</i>	402	4.2
7. <i>perfectly</i>	342	3.6
8. <i>entirely</i>	335	3.5
9. <i>significantly</i>	274	2.9
10. <i>fully</i>	239	2.5
TOTAL	7560	78.7

Table 7: NS: Most frequent amplifiers

Amplifier	NNS occurrences	
	Σ	%
1. <i>really</i>	55	18.2
2. <i>extremely</i>	45	14.9
3. <i>absolutely</i>	32	10.6
4. <i>completely</i>	25	8.3
5. <i>truly</i>	17	5.6
6. <i>totally</i>	16	5.3
7. <i>highly</i>	14	4.6
8. <i>incredibly</i>	14	4.6
9. <i>utterly</i>	13	4.3
10. <i>particularly</i>	10	3.3
TOTAL	241	79.8

Table 8: NNS: Most frequent amplifiers

As we can see, the ten most frequent amplifiers in both databases appeared in the majority of all combinations found in the databases, i.e. almost eighty per cent of all occurrences in both databases were found with these amplifiers. The total per centage of the ten most frequent NS amplifiers is little lower simply due to the higher number of amplifiers in the database. However, these figures support the already mentioned idea relating to the non-native speakers' knowledge and use of collocations, i.e. it seems that these ten most frequent amplifiers all seem to be considered as displaying open collocability and they are used as all-round amplifiers or "safe-bets." In addition, six of the most frequent amplifiers in the NNS database are those which appear in significantly higher number of types. This fact, again, suggests, that the non-native speakers' "creative" ability is much more limited in comparison with the native speakers and more importantly, it further supports the claim stated above, i.e. the non-native speakers seem to use amplifiers more as separate building bricks rather than as parts of prefabricated units (for further evidence see below).

Furthermore, the results of the comparison showed even more important results concerning the differences between the selection and order of native / non-native amplifiers. Except for *really* and *extremely* which were among the three most

frequent amplifiers in both databases, the selection as well as the order of the amplifiers was substantially different. In general, all amplifiers which were common to both groups, i.e. *really*, *extremely*, *completely*, and *totally* + adjective display similar frequencies of occurrences (the only exception is the amplifier *particularly* which demonstrates a very significant level of underuse – see above). Nevertheless, there were five amplifiers which were frequent in one of the databases yet missing in the other: *especially*, *perfectly*, *entirely*, *significantly*, and *fully* as the most frequent amplifiers in the NS database (and not present in the NNS one), and *absolutely*, *truly*, *highly*, *incredibly*, and *utterly* as the most frequent amplifiers in the NNS database (and not present in the NS database). As already mentioned above, four of these five amplifiers demonstrate statistically significant levels of overuse in the NS database. Importantly, these figures, again, showed that there were striking differences between the choices of particular amplifiers in both groups (for details see below). Significantly, these differences revealed one marked tendency on the part of the non-native speakers. If we take into account the amplifiers which are common to both groups (except for *particularly*) together with the amplifiers which are found only on the NNS list, i.e. *really*, *extremely*, *completely*, *totally*, *absolutely*, *truly*, *highly*, *incredibly*, and *utterly*, we found that they either have the direct translation equivalents in Czech or translate very nicely into Czech. On the other hand, the amplifiers found only on the NS list, i.e. *especially*, *perfectly*, *entirely*, *significantly*, and *fully*, either have rather formal translation equivalents, or a relatively much less frequent translation equivalent. Thus, there is a strong L1 influence since the most frequent amplifiers used by the non-native speakers are for the most part congruent with the Czech word combination and presumably result from transfer from Czech.

Another striking difference between the use of collocations, or more particularly between the native-like combinations and those which appeared the most salient on the part of the non-native speakers, can be demonstrated by the comparison of the order of amplifiers used with particular adjectives from the point of view of their frequency, i.e. to what extent the order of the NNS amplifiers corresponded with the order based on the NS database. In other words the more similar the orders, the closer the non-native speakers' knowledge of collocations to the "native-like ideal." In an ideal state the orders concerning particular adjectives would be the same, or with little differences, unfortunately, the contrary seems to be the case.

First, the most frequent NNS amplifier for each adjective was compared with the NS database. Particularly, these most frequent NNS amplifiers were contrasted with the orders of particular NS collocations and the aim of this comparison was to find out whether the most frequent NNS combinations would correspond with the most frequent NS combinations and if not, at which position these NNS amplifiers appeared with their respective adjectives on the NS list. It was decided that if an amplifier was found in first position with the particular adjective in both the NS and NNS database, this amplifier was given 5 points. If it was found in second position with the particular adjective on the NS list, it was given 4 points, and so on. If the most frequent amplifier was not found on the NS list at all, no points were given. Thus, in an ideal state, i.e. if all most frequent NNS amplifiers for each adjective actually corresponded with the most frequent NS amplifiers with the adjective in question, the NNS would get one hundred points (i.e. five points for each adjective). The result of this comparison revealed that the non-native speakers got sixty-eight points (out of one hundred), therefore they approached the sixty-eight per cent of the ideal – which can be considered a rather satisfactory result. In particular, eight most frequent NNS combinations for each adjective were found as the most frequent ones in the NS database, i.e. *completely different*, *extremely difficult*, *really hard*, *absolutely necessary*, *completely reliable*, *really serious*, *highly significant*, and *terribly sorry*. It must be noted that all these amplifiers translate nicely into Czech and are also rather frequent combinations in Czech. Among other most frequent NNS combinations, *really happy*, *extremely important*, and *extremely useful* were the second most frequent choices in the NS database; *extremely bad*, *extremely cold*, and *especially true* occupied the third position on the NS list; and *extremely good*, *absolutely great*, and *really simple* were the fourth on the NS list. *Absolutely miserable* was the only combination found on the fifth position and finally, *highly aware* and *absolutely natural* were not found among the most frequent NS amplifiers of the respective adjectives at all.

If we reversed the comparison, i.e. if we took the most frequent amplifiers for each adjective on the NS list and compared them with the order of the NNS combinations, we found that except for eight amplifiers which were the most frequent in both databases mentioned above, five most frequent NS combinations were found on the second position in the NNS database, that is *really bad*, *really good*, *really great*, *extremely simple*, and *particularly true*. *Fully aware*, *perfectly*

happy, and *particularly important* appeared on the third position in the NNS database; *perfectly natural* occupied the fourth position and importantly, three of the NS combinations of most frequent amplifiers and the respective adjectives in the NS database were not marked by the non-native speakers at all, i.e. *bitterly cold*, *thoroughly miserable*, and *particularly useful*.

In addition, if we compare the orders of amplifiers for the respective adjectives from the point of view of frequency, we found that out of one hundred and eighty-eight combinations, the order, i.e. actual positions of one hundred and twenty-five of them – comprising sixty-six per cent – did not mutually match – in fact it is the sum of types exclusive to the NS and the NNS database (for detail see below). Thus, if we take into account only those combinations that were common to both groups (i.e. eighty six combinations in total), nineteen combinations appeared on the same position in both databases, twenty seven of non-native speakers' combinations were found on higher positions, i.e. they were used more frequently than in the NS database, and seventeen of them were found in the lower positions, i.e. found as less frequent in comparison to the NS database. In particular, thirty-three types appeared on <-1, +1> positions from each other, thirteen types were found on <-2, +2> positions from each other, six types occupied <+3> position from the other and five types were found on <-4, +4> positions from each other.

All in all, there are actually two conclusions that can be made according to these results. First, if we take into account only the most frequent, or the most salient, native collocations, the non-native speakers showed that their knowledge of a significant number of them can be considered satisfactory. However, from the general point of view, the differences between the actual frequencies of both amplifiers and whole word combinations and the orders of the combinations were in most of the cases too large to be ignored. Thus, the general picture is one of learners who seem to use amplifiers more as building bricks rather than as part of prefabricated units, and conspicuously, the L1 influence on the selection and ordering of amplifiers proved to be very strong – yet the L1 influence is in general negative.

So far, it has been proved that there are significant differences between the existing native combinations and those suggested by the non-native speakers. In order to reveal particular, or specific, differences concerning the use of amplifiers by native- and non-native speakers of English, it was necessary to further subdivide the

category of amplifiers into three categories, i.e. the combinations of an amplifier plus and adjective that were exclusively used by native speakers, those that were exclusively used by non-native speakers, and those that were common to both groups (see Table 9):

Database	NS combinations	NS + NNS combinations	NNS combinations
NS database	39 (38%)	63 (62%)	
NNS database		63 (42 %)	86 (58%)

Table 9: Combinations exclusive to native- or non-native speakers and combinations common to both

7.2.2.1. Combinations exclusive to the NS / NNS database

In the NS database, there were 39 different combinations which were not found in the NNS database. These combinations contained twenty-six amplifiers in total and they can be generally considered stereotyped combinations such as *bitterly cold*, *extremely sorry*, *entirely different*, *jolly good*, *keenly aware*, *notoriously difficult*, *painfully hard*, *radically different*, or *vitaly necessary*. Thus, it suggests that this category was problematic for non-native speakers. Indeed, out of the twenty-six amplifiers, fifteen (i.e. thus combinations containing them as well) were not found in the NNS database at all. These fifteen amplifiers were found in seventeen combinations and nine hundred and twenty-one occurrences, i.e. almost ten per cent of all combinations and occurrences respectively were found exclusively in the NS database. These native exclusive combinations were: *apparently simple*, *certainly true*, *clearly necessary*, *desperately hard*, *exceptionally cold*, *extraordinarily difficult*, *genuinely useful*, *jolly good*, *keenly aware*, *notoriously bad*, *notoriously difficult*, *radically different*, *strictly necessary*, *sufficiently serious*, *thoroughly miserable*, *wholly natural*, and *wholly reliable*. As we can see, there were mostly stereotyped combinations that were exclusively used by native speakers. Interestingly, two maximizers also appeared on this list, i.e. *thoroughly*, and *wholly*. Broadly speaking, almost all these amplifiers have no direct translation equivalents, almost none of them translate particularly nicely into Czech and their translation

equivalents can be in most of the cases associated with rather formal Czech. Thus, it can be claimed that the L1 influence, again, plays a prominent part.

The NNS database contained almost sixty per cent of combinations which were used by the non-native speakers yet not occurred among NS combinations consisting of five most frequent amplifiers and their particular adjectives. In particular, there were twenty-seven amplifiers constituting eighty-six different combinations with one hundred and twenty occurrences which were found only in the NNS database. It will be noted that majority of these types, of course, exist in English; however, they cannot be considered as the most typical, or the frequent ones. As already mentioned above (cf. Table 4), there were eleven amplifiers which were found exclusively in the NNS database (i.e. not occurring among the five most frequent amplifiers of respective adjectives): i.e. *actually*, *alarmingly*, *crucially*, *fairly*, *immensely*, *incredibly*, *laughably*, *remarkably*, *simply*, *unbelievably*, and *vastly*. Again, majority of these amplifiers in fact appear with the adjectives in question in the *BNC*, yet they are not among the most frequent ones. Thus, the non-native combinations can be divided into two categories. The first category consists of combinations which actually do exist in English but do not appear among the most frequent ones in the *BNC* (thus in English as well) and the second category contains exclusively non-native speakers combinations which do not exist in English at all.

The first category contains thirty-three combinations containing sixteen amplifiers found in eighty occurrences in total. It contains combinations such as⁹ *absolutely different* (5), *absolutely happy* (3), *actually good* (9), *completely true* (9), *extremely aware* (3), *highly important* (38), *immensely important* (20), *incredibly hard* (9), *terribly good* (19), *totally aware* (5), *truly good* (6), *utterly different* (29), etc. Most of the combinations are stereotyped ones which translate more or less nicely into Czech, thus there is, again, an obvious L1 influence – indeed majority of amplifiers found in these combinations are those demonstrating a very significant level of overuse in the NNS database (cf. Table 5), all of which have either direct translation equivalents or translate very nicely into Czech.

The second category is of paramount interest for it contains combinations suggested by the non-native speakers which, however, are not found in English at all,

⁹ The figures in parentheses following each collocation represent the tokens of particular combinations found in the *BNC*.

i.e. they are exclusively non-native thus erroneous. This category contains fifteen combinations containing ten different amplifiers and eleven adjectives, i.e. *alarmingly bad*, *bitterly true*, *highly good*, *highly serious*, *incredibly natural*, *incredibly serious*, *laughably simple*, *remarkably significant*, *significantly bad*, *significantly important*, *simply miserable*, *truly impossible*, *truly reliable*, *utterly good*, and *utterly great*. As we can see, there is, again, an evident strong influence of L1, i.e. most of these combinations are actually literal translations of very frequent Czech collocations such as *velmi dobrý*, *velmi vážný*, *opravdu nemožný*, *opravdu spolehlivý*, or *naprosto skvělý*. Moreover, the NNS database contained some rare examples of creative combinations such as *alarmingly bad*, *bitterly true* or *laughably simple*. Even though these are not successful collocations in English, they are, again, either literal translations of Czech collocations or they result from transfer from Czech phrases: in Czech we indeed use that something is *směšně jednoduché*, that a situation is becoming *alarmující* (with a negative connotation), or phrase *Je hořkou pravdou...* (which actually exists in English as well, though as a nominal phrase *bitter true*). It must be noted that all these combinations appeared in the NNS database only once except *truly reliable* which was suggested three times. Nevertheless, if we consider that these combinations represent almost ten per cent of the total number of non-native combinations, the number is indeed striking.

7.2.2.2. Combinations common to both NS and NNS database

Overall, there were sixty-three combinations resulting from combinations of twenty amplifiers and twenty adjectives which were found in both databases (i.e. only five most frequent amplifiers of the NS database compared). These combinations contained all amplifiers listed among the most frequent NNS amplifiers (except *utterly*), i.e. those which in general translate well into Czech, together with several examples of the most significant, or prototypical, collocations as well as frequent “creative” collocations, both of which seem to be well-known and used without any difficulties by the non-native speakers. Such collocations found in the both databases were for instance *absolutely necessary*, *acutely aware*, *awfully sorry*, *deadly serious* *entirely happy*, *exceedingly difficult*, *extremely important*, *fully aware*, *highly significant*, *particularly good*, *perfectly happy*, *significantly different*, *terribly different*, *truly great*, *vitaly important*, etc. On the one hand, these findings

suggested that the non-native speakers were aware of the most frequent, i.e. significant, collocation pairs and the truth is that some of the most significant collocations were indeed obviously firmly entrenched in the NNS mental lexicons. On the other hand, we must not forget that the NS database contained “only” five most frequent combinations of an amplifier plus an adjective and, more importantly, there were only sixty-three combinations (out of one hundred and eighty eight combinations in total) which were common to both groups. Thus, the combinations common to both groups represented only thirty-four per cent of the whole database, in other words, only thirty-four per cent of the most frequent English collocations studied in this analysis were salient for the non-native speakers and (to the lesser or greater extent) entrenched in their minds. In general, this number is rather alarming and all these findings further support the claim that the non-native speakers’ knowledge of as well as their ability to use collocations is in general rather weak.

To sum it up, the results of the cloze tests suggest that whereas the non-native speakers may be well aware of some of the most frequent, i.e. prototypical and generally used, English collocations, in general their knowledge of and ability to use collocations proved to be rather weak. One of the most important factors contributing to this state is a strong L1 influence which is presumably responsible for a large number of errors on the part of the non-native speakers. All in all, the non-native speakers’ knowledge of and the ability to use collocations can be hardly considered satisfactory, i.e. they overall did not succeed in attaining the native “ideal.”

7.3. Results of the Significance of Collocation test

7.3.1. Overall results

So far, it has been established that non-native speakers are using collocations which do not always correspond with the majority of the native speakers’ ones, especially with the less frequent, or less salient, ones. In order to further analyse whether the non-native speakers’ sense of salience was really so weak or underdeveloped, Granger’s Significance of collocation test (SOC test) was applied to another group of participants. In contrast to the Cloze test, the SOC test can be considered easier from the point of view of participants, i.e. whereas the

participants in the cloze test had to think up the possible collocations without any hints, the participants in the SOC test had a list of possible combinations at their disposal. Thus, even if the participants in the SOC test were not familiar with particular collocations, it was assumed that the significant, or the most salient, collocations would be striking enough and therefore marked anyway. As already mentioned in Section 6 (Material and Method), the SOC test consisted of eleven amplifiers and the participants were asked to choose the acceptable collocates of these amplifiers from a list of fifteen adjectives in each case. The test was devised in a way that for each amplifier there was usually only one adjective which in combination with the amplifier in question constituted a significant collocation. In some cases there were a few other possible, but not frequent combinations, i.e. there was only one salient collocation for each amplifier. Thus, in an ideal state, the participants should have chosen only those combinations constituting significant collocations, i.e. eleven combinations in total (one for each amplifier).

Overall, if we counted the total number of combinations found in the *BNC* (both the most salient and those which actually exist in English yet are not very frequent) and the number of combinations suggested by the non-native speakers, the comparison of these figures yielded particularly interesting results (see Table 10):

Amplifier	Number of possible combinations - <i>BNC</i>	Number of suggested combinations - <i>NNS</i>
<i>highly</i>	6	14
<i>seriously</i>	3	14
<i>readily</i>	2	12
<i>blissfully</i>	3	14
<i>vitally</i>	5	13
<i>fully</i>	5	13
<i>perfectly</i>	7	14
<i>heavily</i>	2	13
<i>bitterly</i>	2	14
<i>absolutely</i>	9	15
<i>utterly</i>	9	15
TOTAL	53	151

Table 10: Possible combinations (*BNC*) vs. suggested combinations (*NNS*)

As we can see, the non-native speakers suggested one-hundred and fifty-one combinations in comparison to the fifty-three combinations found in the *BNC*. In other words, almost one-hundred of the combinations marked by the non-native speakers were not found in the *BNC* and are therefore regarded as dubious in English. If we take into account that there were one hundred and sixty-five possible combinations overall, the non-native speakers marked almost ninety-two per cent of them. This finding is rather striking, suggesting that the non-native speakers' sense of salience is not only weak, but also partly mistaken. The amplifiers with the highest number of suggested combinations, i.e. the most frequent ones, were *absolutely* (marked in two hundred and forty-three cases) and *highly* (marked in one hundred and eighty-two cases). As already mentioned in the previous sections, *absolutely* has a direct translation equivalent – *absolutně* – and it is one of the most frequent Czech amplifiers, and *highly* – *vysoce* – is also rather frequent Czech amplifier, though it can be considered rather formal; nevertheless, it indeed translates nicely into Czech. Interestingly, there were some amplifiers with no suggested combinations on the part of the non-native speakers, i.e. the amplifier *readily* was left empty eleven times (i.e. in twenty-two per cent of tests), there were seven tests (i.e. fourteen per cent) in which the amplifier *vitally* was not suggested in any combinations, the amplifier *heavily* had no suggested combinations in five tests (i.e. ten per cent), the amplifiers *bitterly* and *fully* appeared in no possible combinations three times and twice respectively. This may suggest that these amplifiers, especially *readily*, *vitally*, and *heavily*, are somewhat problematic from the point of view of the non-native speakers and thus not very well entrenched in their minds as possible amplifiers, i.e. as collocators either.

Further, if we compared the most frequent, i.e. the most salient, combinations of each amplifier and the respective adjective in the *BNC* with those suggested most often by the non-native speakers, there are only minimal differences between these results (see Table 11):

Amplifier	BNC: Most frequent response(s)	NNS: Most frequent response(s)
<i>highly</i>	<i>highly significant</i> (156)	<i>highly significant</i> (42)
	<i>highly important</i> (38)	<i>highly important</i> (37)
<i>seriously</i>	<i>seriously different</i> (227)	<i>seriously different</i> (49)
<i>readily</i>	<i>readily available</i> (426)	<i>readily available</i> (27)
<i>blissfully</i>	<i>blissfully happy</i> (11)	<i>blissfully happy</i> (30)
	<i>blissfully ignorant</i> (6)	<i>blissfully ignorant</i> (30)
<i>vitaly</i>	<i>vitaly important</i> (191)	<i>vitaly important</i> (31)
		<i>vitaly significant</i> (24)
<i>fully</i>	<i>fully aware</i> (239)	<i>fully aware</i> (40)
		<i>fully available</i> (29)
<i>perfectly</i>	<i>perfectly clear</i> (117)	<i>perfectly clear</i> (43)
	<i>perfectly happy</i> (96)	<i>perfectly reliable</i> (20)
<i>heavily</i>	<i>heavily aware</i> (1)	<i>heavily ill</i> (16)
		<i>heavily important</i> (16)
		<i>heavily significant</i> (14)
		<i>heavily difficult</i> (13)
<i>bitterly</i>	<i>bitterly cold</i> (102)	<i>bitterly cold</i> (31)
<i>absolutely</i>	<i>absolutely clear</i> (149)	<i>absolutely impossible</i> (38)
	<i>absolutely essential</i> (122)	<i>absolutely essential</i> (34)
<i>utterly</i>	<i>utterly different</i> (29)	<i>utterly impossible</i> (33)
		<i>utterly different</i> (16)
		<i>utterly ignorant</i> (16)
		<i>utterly miserable</i> (16)

Table 11: Most frequent responses (NS + NNS)

Table 11 shows that there was a good sense of salience among a significant number of the non-native speakers for almost all of the most frequent English collocations. In the case of eight out of eleven amplifiers, the most frequent native combinations corresponded with those suggested by the non-native speakers, indicating that these combinations are in fact very well entrenched in the non-native speakers' mental lexicons. Thus, it seems that these combinations are used not as individual bricks but rather as collocations proper. However, among the most frequent combinations, the

non-native speakers also considered nine other combinations to be significant collocations, though none of them appeared among the most frequent ones in the *BNC*. Particularly, six combinations in fact appeared in the *BNC*, yet these are not the most frequent ones, i.e. *vitally significant*, *fully available*, *perfectly reliable*, *absolutely impossible*, *utterly ignorant* and *utterly miserable*, and curiously, three of the most frequent combinations suggested by the non-native speakers are not accepted in English at all, i.e. *heavily ill / important / difficult*. The case of *heavily* is somewhat paradoxical since this amplifier does not have a direct or clear-cut translation equivalent in Czech. It can be translated as *velice*, or *těžce* and whereas the former fits into all suggested combinations when translated into Czech, it is the latter which is presumably more associated with this amplifier in translation, yet it cannot be used in this particular meaning with suggested adjectives – except, of course, *heavily ill – těžce nemocný*. In this particular case, there is, again, a strong L1 influence.

Thus, the general results have shown two tendencies. First, when it comes to the most frequent and salient English collocations, the non-native speakers showed a remarkable ability to passively recognise (and presumably actively use) these collocations. It seems that these collocations are entrenched in the non-native speakers' mental lexicons and their sense of salience is correct – but only if we take into account these particular combinations. In contrast, there is a second tendency (which is at least as important as the first one) showing that in the case of the other, i.e. less frequent, combinations, the non-native speakers' sense of salience is rather weak and partly mistaken. As already mentioned above, there are ninety-eight combinations suggested by the non-native speakers which do not exist in English. If we take into account that there were one hundred and sixty-five possible combinations in total, almost sixty per cent of the combinations in total suggested by the non-native speakers are not accepted in English. Similarly, if we take into account only the number of combinations suggested by the non-native speakers, almost sixty-five per cent of the combinations marked by the non-native speakers were completely wrong, i.e. non-existing in English. Both these findings suggest that the non-native speakers use a high number of atypical or odd word combinations and again, (except the most frequent collocations), the general picture is one of learners who seem to use amplifiers more as building bricks rather than as part of prefabricated units. Nevertheless, the question is whether these results can be

considered satisfactory or not. On the one hand it can be argued that for the non-native speakers to know the most frequent collocations is enough, i.e. there is no special need for the non-native speakers to know all other possible collocation; yet on the other hand, the lack of knowledge of the other possible – or worse, impossible – collocations can lead to misunderstanding or oddness in discourse or interaction with native speakers of English. In fact, the level of mistakes is too high to be ignored thus in general, the results of the non-native speakers can be said to be far from satisfactory.

7.3.2. Significant collocation

Whereas the previous section compared the native and non-native collocations from the point of view of their frequency, this section focuses on significant collocations, i.e. those combinations which are entrenched in participants' mental lexicons the strongest, thus particularly salient in their minds. As already mentioned in Section 6. (Material and Method), in the SOC tests, the participants were asked not only to choose the acceptable collocates of fifteen adjectives in each case, but also to circle those which in their opinion were more frequently associated with particular amplifiers than all the others, i.e. only one adjective in each case. It was the comparison of the circled forms by the non-native speakers (thus the most salient ones in their minds) and the most frequent native collocations which was relevant for this part of the analysis and it, again, yielded interesting results and showed further differences between the native and non-native use of collocations. For each amplifier, adjectives with the highest significant frequencies of co-occurrence in the *BNC* were taken into account for the comparison. In most of the cases, there were only one or two adjectives showing a significant level of co-occurrence and these were compared to combinations circled by the non-native speakers see Table 12):

Amplifier	Number of significant complementations - <i>BNC</i>	Number of circled complementations - <i>NNS</i>
<i>highly</i>	2	8
<i>seriously</i>	1	5
<i>readily</i>	1	6
<i>blissfully</i>	2	5
<i>vitaly</i>	1	6
<i>fully</i>	1	5
<i>perfectly</i>	2	6
<i>heavily</i>	2	10
<i>bitterly</i>	1	6
<i>absolutely</i>	2	8
<i>utterly</i>	3	9
TOTAL	20	74

Table 12: The number of significant collocations in the *BNC* and the *NNS* database

As we can see, the non-native speakers circled seventy-four combinations in contrast to twenty combinations with significant levels of co-occurrence extracted from the *BNC*. Thus, it means that there were fifty-four more combinations which were particularly salient in the non-native speakers' minds than the most frequently co-occurring combinations found in the *BNC*. In other words, Table 12 gives clear evidence of the non-native speakers' misguided sense of salience. In contrast to Granger's study in which the non-native speakers marked over one hundred fewer combinations constituting significant collocations (from their point of view) than the native speakers, thus their sense of salience was weak; in this case the sense of salience of the non-native speakers was greatly exaggerated. Nevertheless, it must not be forgotten that Granger's study was based on a different kind of data, hence the differences between the number of responses on the part of native speakers. Yet it actually does not make a difference whether the non-native speakers marked significantly lower or higher number of what in their opinion constituted significant collocations – both variants showed the weak as well as mistaken sense of salience on the side of the non-native speakers.

If we take a closer look at the combinations circled by the non-native speakers and compare them with the most frequently occurring combinations in the *BNC*, we also get particularly interesting results revealing specific differences between the significant collocations extracted from the *BNC* and those which were considered significant by the non-native speakers (see Table 13):

Amplifiers	Native-speaker responses	Learner responses
highly	<i>highly significant</i> (156) <i>highly important</i> (38)	<i>highly important</i> (16) <i>highly significant</i> (13) <i>highly reliable</i> (6) <i>highly aware</i> (2) <i>highly essential</i> (2)* <i>highly impossible</i> (2)* <i>highly available</i> (1)* <i>highly ignorant</i> (1)*
seriously	<i>seriously ill</i> (227)	<i>seriously ill</i> (45) <i>seriously cold</i> (1)* <i>seriously difficult</i> (1)* <i>seriously important</i> (1) <i>seriously impossible</i> (1)*
readily	<i>readily available</i> (426)	<i>readily available</i> (26) <i>readily aware</i> (5) <i>readily difficult</i> (1)* <i>readily essential</i> (1)* <i>readily reliable</i> (1)* <i>readily significant</i> (1)*
blissfully	<i>blissfully happy</i> (11) <i>blissfully ignorant</i> (6)	<i>blissfully ignorant</i> (20) <i>blissfully happy</i> (6) <i>blissfully aware</i> (4)* <i>blissfully clear</i> (2) <i>blissfully cold</i> (2)*
vitality	<i>vitality important</i> (191)	<i>vitality important</i> (24) <i>vitality essential</i> (7) <i>vitality significant</i> (5) <i>vitality aware</i> (1) <i>vitality clear</i> (1)* <i>vitality impossible</i> (1)*
fully	<i>fully aware</i> (239)	<i>fully aware</i> (25) <i>fully available</i> (8) <i>fully reliable</i> (4) <i>fully clear</i> (3) <i>fully impossible</i> (1)*
perfectly	<i>perfectly clear</i> (117) <i>perfectly happy</i> (96)	<i>perfectly clear</i> (34) <i>perfectly aware</i> (3) <i>perfectly happy</i> (3) <i>perfectly reliable</i> (3) <i>perfectly different</i> (2) <i>perfectly available</i> (1)
heavily	<i>heavily aware</i> (1) <i>heavily significant</i> (1)	<i>heavily ill</i> (13)* <i>heavily important</i> (7)*

		<i>heavily significant (6)</i> <i>heavily ignorant (3)*</i> <i>heavily essential (2)*</i> <i>heavily aware (1)</i> <i>heavily cold (1)*</i> <i>heavily difficult (1)*</i> <i>heavily miserable (1)*</i> <i>heavily reliable (1)*</i>
<i>bitterly</i>	<i>bitterly cold (102)</i>	<i>bitterly cold (28)</i> <i>bitterly aware (4)</i> <i>bitterly miserable (4)*</i> <i>bitterly ignorant (2)*</i> <i>bitterly happy (1)*</i> <i>bitterly impossible (1)*</i>
<i>absolutely</i>	<i>absolutely clear (149)</i> <i>absolutely essential (122)</i>	<i>absolutely clear (13)</i> <i>absolutely impossible (12)</i> <i>absolutely essential (9)</i> <i>absolutely important (2)</i> <i>absolutely aware (1)*</i> <i>absolutely reliable (1)</i> <i>absolutely significant (1)*</i>
<i>utterly</i>	<i>utterly different (29)</i> <i>utterly impossible (11)</i> <i>utterly miserable (10)</i>	<i>utterly impossible (19)</i> <i>utterly happy (6)</i> <i>utterly miserable (6)</i> <i>utterly different (4)</i> <i>utterly ignorant (3)</i> <i>utterly clear (1)</i> <i>utterly essential (1)</i> <i>utterly important (1)*</i>

Table 13: Significant collocations: the *BNC* vs. the *NNS* database

Note: Combinations suggested by the non-native speakers which are atypical, i.e. do not occur in the *BNC*, are indicated by an asterisk.

As we can see in Table 13 there were, again, two findings supporting the general claim made above. First, Table 13 gives clear evidence that the non-native speakers had a good sense of salience for some combinations, in particular those with the highest frequency of occurrence in the *BNC*. From the point of view of the combinations which were the most salient in the participant's minds, in the case of seven amplifiers the non-native speakers indeed most often circled those combinations which were extracted from the *BNC* as those with the highest level of occurrence. This fact supports the claim that the non-native speakers have a great sense of salience concerning the most typical, or prototypical, English collocations. Furthermore, in the case of four amplifiers where the non-native speakers considered to be most salient combinations other than those with the most frequent co-

occurrence in the *BNC*, in each case the most salient English collocation actually appeared among the circled ones by the non-native speakers. Particularly, in the case of *highly* and *blissfully*, the two combinations with the highest level of co-occurrence in the *BNC* appeared in the reverse order in the non-native speakers' responses. Interestingly, the second combination extracted from the *BNC* in each case displayed a high level of occurrence, yet in comparison with the most frequent combinations in each case, both second combinations appeared significantly less often. However, in the non-native speakers' responses, the first option was circled significantly more often than the second one (which is in fact the most frequently occurring combination in the *BNC*) in each case. On the one hand, this finding again supports the claim concerning a good sense of salience among the non-native speakers because they selected the two most frequent English combinations. Yet on the other hand, since the order of these two combinations is reversed, it may be little indication that their sense of salience was partly mistaken. Concerning the amplifiers *utterly* and *heavily*, they were actually the only two amplifiers in which the order of combinations based on the frequency was remarkably different between the *BNC* and the non-native speakers' data. Whereas in the *BNC* the combination with the highest frequency of occurrence was *utterly different* with twenty-nine occurrences, in the NNS database this combination occupied the third place with only four cases in which this particular combination was circled. The most salient combination of the non-native speakers was *utterly impossible* circled in nineteen cases followed by *utterly happy* and *utterly miserable* circled six times in each case. The case of *heavily* is, again, the most complicated. Whereas there are only two possible adjectives complementing this amplifier in the *BNC* and each of them appeared in the *BNC* only once (thus it can be hardly considered one of the typical, or usual amplifier), the non-native speakers circled eight other combinations as the salient ones. Interestingly, *heavily ill* was circled thirteen times, i.e. this combination represented the most salient collocation for more than one third of the participants, and almost twenty per cent of participants circled the combination *heavily important*. It has already been explained that *heavily ill* operates on the basis of the L1 influence, representing a translation equivalent of the Czech phrase *těžce nemocný*, yet, there is no similar explanation for the combination *heavily important* and the question remains why this particular combination is so deeply entrenched in the non-native speakers' mental lexicons.

The second finding from the comparison of data in Table 13 is related to combinations which were at the same time salient in the non-native speakers' minds yet not found among the combinations with the highest level of occurrence in the *BNC*. As already mentioned above, even though there was evidence of a good sense of salience among a significant number of the non-native speakers for the combinations occurring most frequently in the *BNC*, the same cannot be claimed from the point of view of other combinations suggested by the non-native speakers. The results of the SOC test showed that there was a significantly higher number of combinations considered salient on the part of the non-native speakers which were not found among the salient ones extracted from the corpus. Again, it was necessary to divide these non-native speakers' combinations into two groups, i.e. those which actually exist in English yet cannot be considered frequent, i.e. typical; and those which are completely unnatural, i.e. not existing in English. Thus, apart from the eighteen combinations with the most frequent occurrence suggested by the non-native speakers (more or less corresponding with those found in the *BNC*), the NNS database contained twenty three combinations which were considered most salient by the non-native speakers, yet not found among the frequent combinations in the *BNC*. Furthermore and crucially, there were thirty-three combinations marked as the most salient by the non-native speakers which, however, do not exist in English at all. In other words, if we take into account that the NNS database contained the total of seventy-four circled combinations, fifty-five per cent of the non-native speakers' combinations in fact exist in English. However, more than half of these combinations are not typical, i.e. they are rather infrequent combinations in English. What is even more striking, forty-five per cent of the combinations considered most salient by the non-native speakers actually do not seem to exist in English at all, i.e. they are completely incorrect.

To sum it up, the results of the SOC test were rather similar to those of the cloze tests. The comparison of the native and the non-native use of collocations revealed significant differences between these two groups. The knowledge and ability to use collocations on the part of the non-native speakers were not entirely convincing. It is true that the non-native speakers showed a good sense of salience in the most frequent or the most salient collocations; however, in other cases, i.e. English collocations which are not the most significant, the level of error was even higher in the SOC test than in the cloze tests. Thus, the number of mistakes relating

to less frequent combinations is, again, too high to be ignored, supporting the claim that the non-native speakers knowledge of and ability to use collocations is rather weak, their sense of salience is significantly mistaken (in most of the cases), thus it further supports the general picture of the non-native speakers using amplifiers as individual bricks rather than as parts of prefabricated units.

7.4. Pedagogical implications

The results of the present study have demonstrated that the non-native speakers' phraseological skills are severely limited. On the one hand, the non-native speakers indeed use native-like prefabricated units, yet the number of these is relatively low. On the other hand, the non-native speakers totally fail concerning the majority of English prefabricated units, especially those not very frequent or creative, and, most importantly, the non-native speakers produce a vast number of atypical, foreign-sounding combinations which do not exist in English. Thus, it is vitally important to lay greater emphasis on prefabricated units – or the prefabricated nature of language in general – in ELT. These limitations on the part of the non-native speakers are generally known and accepted; however, as Granger points out, we actually know very little about prefabricated units. We still do not have enough information concerning the acquisition of prefabricated units, what difficulties they cause in production, and how the interaction between L1 and L2 prefabricated units actually works. Moreover, it is still not clear what role prefabricated units should play in L2 teaching, particularly what and how much to teach, or how to teach them in general (cf. Granger, 1998: 159).

In my opinion, it is – first and foremost – absolutely necessary to explain the prefabricated nature of the English lexicon (or that of any other language in general) to non-native speakers. From my own experience, the learners have indeed very limited notion that the internal structure of a language largely consists of prefabricated lexical units rather than individual bricks which can be put together without restrictions (except, of course, in grammar). It is indeed necessary to make learners aware of this fact and thus improve their strategies of language acquisition. If, and only if, learners start to be aware of language patterning and take it seriously into account while learning, can there be a chance or a possibility of improving their

knowledge of prefabricated units thus of improving their ability to produce native-like combinations. Second, it is also crucial to explain to learners that they cannot rely on Czech when producing English word-combinations, clauses or sentences, or utterances. They indeed have to become aware of the fact that there are vast differences on all levels between Czech and English (or any other language) and that it is not possible – in most of the cases – to build English constructions upon the Czech model, i.e. there is rarely a one-to-one relationship between the structures of the two languages. In other words, non-native speakers have to understand that their mother language inevitably plays a prominent role, yet this role is in the majority of cases rather negative, resulting in atypical, strange sounding, or incorrect combinations in English.

However, these two steps should be applied to language acquisition in general, rather than be considered exclusive to the prefabricated nature of language. If we focus on prefabricated units only, it is indeed hard to make any conclusions or suggestions as to how to improve learners' knowledge of these units and ability to use them. Since it has been shown that exposure to L2 does not significantly improve this knowledge, the development of new EFL teaching methods and materials which would lay greater emphasis on prefabricated units seems inevitable – though it will be presumable long and difficult process from the point of view of problems which have to be solved prior to developing these new methods and materials). Nevertheless, the general advice applicable to present-day ELT is to concentrate not only on teaching learners how to produce grammatically correct constructions, but also on giving more prominence to contexts, or more particularly to co-texts, in which these combinations appear as well as to paradigmatic and syntagmatic relations in general. In my opinion, it is necessary to increase learners' exposure to naturally occurring English (for instance by greater emphasis on reading English texts or watching English films, series, etc.) while at the same time making them alert to the phraseological aspects of language use. This is probably the way to a better acquisition of various constructions (including prefabricated units) and maybe to a lower number of errors learners produce when using English.

8. Conclusion

The initial hypothesis for this study was that learners' knowledge of and ability to use collocations would be much weaker in comparison with native speaker data. In particular, it was presupposed that learners would make less use of native-like collocations than native speakers except for a small number of most frequent collocations which would be underused in the investigated data (and thus of course in their actual use of English, both spoken and written), that the preferred amplifiers would differ in comparison with native speaker data, and that learners would make a number of mistakes in their use of collocations. The results of both tests, i.e. the Cloze test as well as the Significance of collocation test confirmed (to a lesser or greater extent) the initial hypotheses.

In the Cloze test, the overall number of amplifiers analysed was forty eight. These amplifiers were preceded by twenty different adjectives, resulting in one hundred and eighty-eight different combinations in total (this resulting number of amplifiers combined amplifiers suggested by the non-native speakers in cloze tests and five most frequent amplifiers for the respective adjectives extracted from the *BNC*). Since only five most frequent amplifiers for each adjective were extracted from the *BNC*, there was no significant underuse of the NNS amplifiers in the present data as expected (nevertheless, if we compared the NNS results with all possible amplifiers for each adjective, there would indeed be a significant underuse of amplifiers in the NNS database). However, the mere comparison of the combinations suggested by the non-native speakers and those extracted from the *BNC* revealed first substantial differences in the use of collocations by native- and non-native speakers of English. First, the results revealed striking differences between the use or selection of amplifiers in the NS and the NNS database. In particular, there were eleven amplifiers used by the non-native speakers which were not found among five most frequent amplifiers of the NS database. Similarly, there were fifteen amplifiers which were found among the most frequent ones in the NS database yet not suggested by the non-native speakers. Second, the overall results showed the statistically significant difference between the existing native combinations and those suggested by the non-native speakers, i.e. whereas there were one-hundred and two (most frequent) combinations in the NS database in total, the non-native speakers suggested one-hundred and forty-nine combinations. In addition, the non-native speakers generally

used the same amplifiers regardless of the following adjective, revealing not only that their sense of salience was somewhat weaker, but also that their creative ability was significantly limited in comparison with native-speakers' data. Among the amplifiers which were overused by the non-native speakers, *absolutely* (having a direct translation equivalent in Czech), *truly*, and *incredibly* (both translating very nicely into Czech) demonstrated very significant levels of overuse on the part of the non-native speakers. Importantly, in the case of all of the amplifiers found with the highest number of combinations suggested by the non-native speakers, there was an obvious and strong L1 influence, i.e. the amplifiers with direct translation equivalents or those which translate nicely into Czech were the most frequent ones in the NNS database, as if there were no differences between English and Czech, i.e. as if there was a direct, straight way (one-to-one relationship) of translating these combinations from English to Czech or vice versa.

The next step was to find out to what extent – or if at all – the non-native speakers attained the native “ideal,” i.e. how the selections as well as orders of particular combinations suggested by them corresponded with the existing native ones. The comparison of the most frequent amplifiers in both databases showed that the selection as well as the order of the amplifiers by native- and non-native speakers were substantially different. Except for four amplifiers which were common to both groups, i.e. *really*, *extremely*, *completely*, and *totally* (which displayed similar frequencies of occurrences), there were five amplifiers which were frequent in one of the databases yet missing in the other: *especially*, *perfectly*, *entirely*, *significantly*, and *fully* in the NS database (and not present in the NNS one), and *absolutely*, *truly*, *highly*, *incredibly*, and *utterly* in the NNS database (and not present in the NS database). Significantly, these differences revealed one marked tendency on the part of the non-native speakers. If we took into account the amplifiers which were common to both groups together with the amplifiers which were found only on the NNS list, i.e. *really*, *extremely*, *completely*, *totally*, *absolutely*, *truly*, *highly*, *incredibly*, and *utterly*, we found that they either had the direct translation equivalents in Czech or translated very nicely into Czech. On the other hand, the amplifiers found only on the NS list, i.e. *especially*, *perfectly*, *entirely*, *significantly*, and *fully*, either had rather formal translation equivalents or a relatively much less frequent translation equivalent. Thus, there was a strong L1 influence since the most frequent amplifiers used by the non-native speakers were for the most part congruent with the Czech word combination

and presumably resulted from transfer from Czech. The comparison of the orders of amplifiers used with particular adjectives from the point of view of their frequency (i.e. to what extent the order of the NNS amplifiers corresponded with the order based on the NS database) revealed two conclusions. First, if only the most frequent, or the most salient, native collocations were taken into account, the non-native speakers showed that their knowledge of a significant number of them could be considered satisfactory. However, from the general point of view, the differences between the actual frequencies of both amplifiers and whole word combinations and the orders of the combinations were in most of the cases too large to be ignored. Thus, the general picture was one of learners who seemed to use amplifiers more as building bricks rather than as part of prefabricated units, and conspicuously, the L1 influence on the selection and ordering of amplifiers proved to be very strong – yet the L1 influence was in general negative.

In order to reveal particular, or specific, differences concerning the use of amplifiers by native- and non-native speakers of English, it was necessary to further subdivide the category of amplifiers into three categories, i.e. the combinations of an amplifier plus an adjective that were exclusively used by native speakers, those that were exclusively used by non-native speakers, and those that were common to both groups. Broadly speaking, the combinations exclusive to the NS database could be generally considered stereotyped combinations and almost all amplifiers found in these combinations had no direct translation equivalents and almost none of them translated particularly nicely into Czech (in fact, their translation equivalents could be in most of the cases associated with rather formal Czech). Thus, it seemed that the L1 influence, again, played a prominent part. The non-native exclusive combinations were divided into two categories, i.e. those which actually existed in English but did not appear among the most frequent ones in the *BNC* (thus in English as well) and those which did not exist in English at all. The first category contained thirty-three combinations and, again, most of the combinations were stereotyped ones which translated more or less nicely into Czech. The second category was of paramount interest for it contained fifteen combinations (mainly stereotyped ones though there were also rare examples of creative combinations) which were exclusively non-native but which were not found in the *BNC* at all. All in all, there was, once again, an evident strong influence of L1, i.e. most of these combinations were actually literal translations of very frequent Czech collocations such as *velmi dobrý*, *velmi vážný*,

opravdu nemožný, opravdu spolehlivý, or naprosto skvělý. Moreover, the creative combinations (which in fact were not successful ones in English) such as *alarmingly bad, bitterly true* or *laughably simple* were either literal translations of Czech collocations or they resulted from transfer from Czech phrases. Though all these combinations appeared in the NNS database only once (except for *truly reliable* which was suggested three times) they represented almost ten per cent of the total number of non-native combinations, and this number was indeed striking. From the point of view of combinations common to both NS and NNS database, these combinations either contained amplifiers which in general translated well into Czech or could be included in the category of the most significant, or prototypical, frequent English collocations (which seemed to be well-known and used without any difficulties by the non-native speakers). On the one hand, these findings suggested that the non-native speakers were aware of the most frequent, i.e. significant, collocation pairs and the truth was that some of the most significant collocations indeed seemed obviously firmly entrenched in the NNS mental lexicons. On the other hand, the combinations common to both groups represented only thirty-four per cent of the whole database, in other words, only thirty-four per cent of the most frequent English collocations studied in this analysis were salient for the non-native speakers and (to the lesser or greater extent) entrenched in their minds. These findings further supported the claim that the non-native speakers' knowledge of as well as their ability to use collocations is in general rather weak. Importantly, a strong L1 influence was presumably responsible for a large number of errors on the part of the non-native speakers. All in all, the results of the cloze tests suggested that the non-native speakers' knowledge of and the ability to use collocations could be hardly considered satisfactory, i.e. they overall did not succeed in attaining the native "ideal."

In order to further analyse whether the non-native speakers' sense of salience was really so weak or underdeveloped, Granger's Significance of collocation test (SOC test) was applied to another group of participants. In contrast to the Cloze test, the SOC test could be considered easier from the point of view of participants for (unlike in the Cloze test) they had a list of possible combinations at their disposal. In other words, it was assumed that even if they were not familiar with particular collocations, the significant, or the most salient, ones would be striking enough and therefore marked anyway. However, the results of the SOC test were, unfortunately, similar to those of the cloze tests. In the SOC test, the non-native speakers suggested

one-hundred and fifty-one combinations in contrast to fifty-three combinations found in the *BNC*. In other words, almost one-hundred of the combinations marked as significant by the non-native speakers were not found in the *BNC* and are therefore regarded as dubious in English, suggesting that the non-native speakers' sense of salience was not only weak, but also partly mistaken. Again, the amplifiers with the highest number of suggested combinations, i.e. the most frequent ones, were *absolutely* (marked in two hundred and forty-three cases) and *highly* (marked in one hundred and eighty-two cases). As already mentioned in the previous sections, *absolutely* had a direct translation equivalent – *absolutně* – and it was one of the most frequent Czech amplifiers, and *highly* – *vysoce* – was also rather frequent Czech amplifier, though it could be considered rather formal; nevertheless, it indeed translated nicely into Czech. Interestingly, there were several amplifiers with no suggested combinations on the part of a number of the non-native speakers, i.e. *readily*, *vitally*, *heavily*, *bitterly* and *fully*. This suggested that these amplifiers were somewhat problematic from the point of view of the non-native speakers and thus not very well entrenched in their minds as possible amplifiers, i.e. as collocators either. Nevertheless, from the point of view of the most frequent combinations of each amplifier and the respective adjective in the *BNC* and those suggested most often by the non-native speakers, the results showed only minimal differences. It suggested that there was a good sense of salience among a significant number of the non-native speakers for almost all of the most frequent English collocations, indicating that these combinations were in fact very well entrenched in the non-native speakers' mental lexicons and thus that they were used not as individual bricks but rather as collocations proper.

Thus, the results revealed, again, two general tendencies. First, when it came to the most frequent and salient English collocations, the non-native speakers showed a remarkable ability to passively recognise (and presumably actively use) these collocations. In contrast, there was a second tendency showing that in the case of the other, i.e. less frequent, combinations, the non-native speakers' sense of salience was rather weak and partly mistaken. In particular, there were ninety-eight combinations in the database suggested by the non-native speakers which did not exist in English. In other words, if we took into account that there were one hundred and sixty-five possible combinations in total, almost sixty per cent of the combinations suggested by the non-native speakers were not accepted in English (similarly, if we

took into account only the number of combinations suggested by the non-native speakers, almost sixty-five per cent of the combinations marked by the non-native speakers were completely wrong, i.e. non-existing in English). To sum it up, the non-native speakers used a high number of atypical or odd word combinations and again, (except the most frequent collocations), the general picture was one of learners who seemed to use amplifiers more as building bricks rather than as part of prefabricated units. In fact, the level of mistakes was too high to be ignored thus in general, the results of the non-native speakers could be said to be far from satisfactory.

The last part of the analysis focused solely on significant collocations, i.e. those combinations which were entrenched in participants' mental lexicons the strongest, thus particularly salient in their minds. It was the comparison of the circled forms by the non-native speakers (thus those regarded significant) and the most frequent native collocations which was relevant for this part of the analysis and it, again, yielded interesting results and showed further differences between the native and non-native use of collocations. Once again, there were huge differences between the combinations considered significant on the part of the non native speakers and significant collocations (i.e. combinations with the most significant levels of co-occurrence) extracted from the *BNC*. In particular, there were fifty-four more combinations which were particularly salient in the non-native speakers' minds than the most frequently co-occurring combinations found in the *BNC*, providing further evidence of the non-native speakers' misguided sense of salience. Interestingly, in contrast to Granger's study in which the non-native speakers marked over one hundred fewer combinations constituting significant collocations (from their point of view) than the native speakers, thus their sense of salience was weak; in this case the sense of salience of the non-native speakers was greatly exaggerated (nevertheless, it must not be forgotten that Granger's study was based on a different kind of data, hence the differences between the number of responses on the part of native speakers). Either way, both variants showed the weak as well as mistaken sense of salience on the side of the non-native speakers. A closer look at the combinations circled by the non-native speakers and at the most frequently occurring combinations in the *BNC* showed two findings supporting the general claim made above. On the one hand, the non-native speakers displayed a great sense of salience concerning the most typical, or prototypical, English collocations. On the other hand, the same cannot be claimed from the point of view of other combinations suggested by the non-native speakers.

The results of the SOC test showed that there was a significantly higher number of combinations considered salient on the part of the non-native speakers which were not found among the salient ones extracted from the corpus. In particular, there were twenty three combinations (actually existing in English) which were considered most salient by the non-native speakers yet at the same time were not found among the most frequent combinations in the *BNC*. Yet crucially, there were thirty-three combinations marked as the most salient by the non-native speakers which, however, did not exist in English at all. In other words, if we took into account that the NNS database contained the total of seventy-four circled combinations, fifty-five per cent of the non-native speakers' combinations in fact existed in English. However, more than half of these combinations were not typical, i.e. they were rather infrequent combinations in English. What was even more striking, forty-five per cent of the combinations considered most salient by the non-native speakers actually did not seem to exist in English at all, i.e. they were completely incorrect. As already mentioned above, the results of the SOC test were rather similar to those of the cloze tests. To sum it up, the knowledge and ability to use collocations on the part of the non-native speakers were not entirely convincing (in fact, the level error in the case of English collocations which were not the most significant was even higher in the SOC tests than in the cloze tests). All in all, the number of mistakes relating to less frequent combinations was, again, too high to be ignored, supporting the claim that the non-native speakers knowledge of and ability to use collocations was rather weak, their sense of salience was significantly mistaken (in most of the cases), thus it further supported the general picture of the non-native speakers using amplifiers as individual bricks rather than as parts of prefabricated units.

To conclude, the results of the present study have confirmed the results of Granger's study, i.e. they have demonstrated that the non-native speakers' phraseological skills are severely limited. On the one hand, the non-native speakers indeed used native-like prefabricated units, yet the number of these was relatively low. On the other hand, the non-native speakers totally failed concerning the majority of English collocations, especially those not very frequent or creative ones, and, most importantly, the non-native speakers produced a vast number of atypical, foreign-sounding combinations which did not exist in English.

References

- Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In Cowie, A. P. (ed.) *Phraseology. Theory, Analysis, and Applications*, 101-122. Oxford: Oxford University Press.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman Grammar of Spoken and Written English*. Longman: Harlow.
- Burger, H. (1998). *Phraseologie. Eine Einführung am Beispiel des Deutschen*. Berlin: Erich Schmidt.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris
- Cowie, A. P. (1981). The treatment of collocations and idioms in learners' dictionaries. *Applied Linguistics* 2 (3): 223–235.
- Cowie, A.P. (1994). Phraseology. In Asher, R.E. (ed.). *The Encyclopedia of Language and Linguistics*, 3168-3171. Oxford: Oxford University Press.
- Cowie, A. P. (ed.) (1998). *Phraseology. Theory, Analysis, and Applications*, 101-122. Oxford: Oxford University Press.
- De Cock, S. (2003). Recurrent Sequences of Words in Native Speaker and Advanced Learner Spoken and Written English. PhD dissertation, Université catholique de Louvain.
- Dušková, L., et al. (1988), *Mluvnice současné angličtiny na pozadí češtiny*, Prague, Academia. 3rd ed. 2003.
- Ellis, N. C. (2008). Phraseology: The periphery and the heart of language. In Meunier, F. & S. Granger (eds.) *Phraseology in Foreign Language Learning and Teaching*, 1-13. Amsterdam: John Benjamins.
- Erman, B. & B. Warren (2000). The idiom principle and the open choice principle. *Text*, 20, 29-62.
- Firth, J. R. (1957). *A synopsis of linguistic theory, 1930–1955*. In *Studies in Linguistic Analysis, Vol. 1952-59*, 1-12. Oxford: The Philological Society.
- Fries, C.C. (1952). *The Structure of English*. New York, NY: Harcourt, Brace and Co.
- Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: The University of Chicago Press.
- Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Granger, S. (1998). Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae. In Cowie, A. P. (ed.) *Phraseology. Theory, Analysis, and Applications*, 145-160. Oxford: Oxford University Press.
- Granger, S. & F. Meunier (eds.) (2008). *Phraseology: An Interdisciplinary Perspective*. Amsterdam: John Benjamins.
- Granger, S. & F. Meunier (2008). Introduction: The many faces of phraseology. In Granger, S. & F. Meunier (eds.) *Phraseology: An Interdisciplinary Perspective*, xix-xxviii. Amsterdam: John Benjamins.

- Granger, S. & M. Paquot (2008). Disentangling the phraseological web. In Granger, S. & F. Meunier (eds.) *Phraseology: An Interdisciplinary Perspective*, 27-49. Amsterdam: John Benjamins.
- Greenbaum, S. (1970). *Verb-intensifier collocations in English*. The Hague: Mouton.
- Gries, S. T. (2008). Phraseology and linguistic theory: A brief survey. In Granger, S. & F. Meunier (eds.) *Phraseology: An Interdisciplinary Perspective*, 3-25. Amsterdam: John Benjamins.
- Handl, S. (2008). Essential collocations for learners of English: The role of collocational direction and weight. In Meunier, F. & S. Granger (eds.) *Phraseology in Foreign Language Learning and Teaching*, 43-66. Amsterdam: John Benjamins.
- Harris, Z. (1982). *A grammar of English on mathematical principles*. New York, NY: John Wiley and Sons.
- Harris, Z. (1991). *A theory of language and information: A mathematical approach*. Oxford: Oxford University Press.
- Hausmann, F.J. (1997). Semiotaxis und Wörterbuch. In Konecny, K.P. & A. Lehr (eds.) *Linguistische Theorie und lexikographische Praxis*, 19-23. Tübingen: Niemeyer.
- Herbst, T. (1996). What are collocations: Sandy beaches or false teeth? *English Studies*, 77 (4), 379-393.
- Hunston, S. & G. Francis (2000). *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English*. Amsterdam: John Benjamins.
- Jones, S. & J. Sinclair, (1974). English lexical collocations. In *Cahiers de Lexicologie*, 24, 15-61.
- Kjellmer, G. (1987). Aspects of English collocations. In Meijs, W. (ed.) *Corpus linguistics and beyond*, 133-140. Amsterdam: Rodopi.
- Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites*. Stanford, CA: Stanford University Press.
- Leech, G., Rayson, P., Wilson, A. (2001). *Word Frequencies in Written and Spoken English* (based on the BNC). Harlow: Longman.
- Lipka, L. (2002). *English Lexicology*. Tübingen: Niemeyer.
- Mel'čuk, I. (1995). Phrasemes in language and phraseology in linguistics. In Everaert, M., E. J. Van der Linden & A. Schenk (eds.) *Idioms: Structural and Psychological Perspectives*, 167-232. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Mel'čuk, I. (1998). Collocations and lexical functions. In Cowie, A. P. (ed.) *Phraseology. Theory, Analysis, and Applications*, 22-53. Oxford: Oxford University Press.
- Meunier, F. & S. Granger (eds.) (2008). *Phraseology in Foreign Language Learning and Teaching*. Amsterdam: John Benjamins.
- Nesselhauf, N. (2004a). What are collocations? In Allerton, D, N. Nesselhauf & Paul Skandera (eds.) *Phraseological Units: Basic Concepts and Their Application*, 1-21. Basel: Schwabe.

- Nesselhauf, N. (2005). *Collocations in a Learner Corpus*. Amsterdam: John Benjamins.
- Palmer, F. R. (1968). *Selected papers of J.R. Firth, 1952–59*. London: Longman.
- Pawley, A. & F. H. Syder (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Richards, J.C. & R. W. Schmidt (eds.) *Language and Communication*, 191–226. London: Longman.
- Quirk et al. (1985) *A Comprehensive Grammar of the English Language*. Longman: Harlow.
- Scott, M. & C. Tribble (2006). *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins.
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sinclair, J. (2008). Preface. In Granger, S. & F. Meunier (eds.) *Phraseology: An Interdisciplinary Perspective*, xv-xviii. Amsterdam: John Benjamins.
- Stubbs, M. (2002). Two quantitative methods of studying phraseology in English. *International Journal of Corpus Linguistics* 7 (2), 215-244.
- Stubbs, M. (2007). An example of frequent English phraseology: Distribution, structures and functions. In Facchinetti, R. (ed.) *Corpus Linguistics 25 Years on*, 89-105. Amsterdam: Rodopi.
- Vachek, J. (2005). *Lingvistický slovník Pražské školy*. Praha: Karolinum.

Corpora

The British National Corpus (BYU-BNC). URL: <http://corpus.byu.edu/bnc/>
Český národní korpus. URL: <http://korpus.cz/>

Dictionaries

Cambridge Advanced Learner's Dictionary 2nd Edition, CD-ROM, (2006).

České resumé

Tato práce je zaměřena na testování anglických kolokací u českých studentů angličtiny. Má za cíl jednak popsat a prozkoumat kolokace v angličtině se zaměřením na různé přístupy ke kolokacím obecně (a z nich plynoucí definice a klasifikace pojmu „kolokace“), a také nastínit znalosti českých studentů angličtiny jak (pasivního) rozpoznání, tak (aktivního) používání anglických kolokací. Práce je založena na studii Sylvianne Granger (1998) „Prefabricated Patterns In Advanced EFL Writing: Collocations and Formulae, v níž popisuje a zkoumá anglické kolokace a formule z pohledu francouzských studentů angličtiny. Testování anglických kolokací u českých studentů angličtiny bylo provedeno na základě dvou typů testů, z nichž každý byl použit v jiné skupině tazatelů. Skupiny se od sebe odlišovaly nejen počtem tazatelů, ale hlavně věkem a dobou studia angličtiny, jakožto druhého jazyka.

V teoretické části práce je nejprve představen pojem „souslovní“ / „ustálená víceslovné pojmenování“ / „ustálená slovní spojení“ (anglický termín „prefabricated units“) a to v souvislosti s frazeologií, která se teprve nedávno dostala do popředí zájmu a stala se jednou z nejdůležitějších lingvistických disciplín. Kapitola 2 se zabývá frazeologií a frazeologickými jednotkami především z obecného hlediska. Podkapitola 2.1. v první řadě definuje pojem frazeologie a popisuje základní princip této disciplíny. Frazeologie se definuje jako věda zabývající se slovními spojeními (spíše než jednotlivými slovy), konkrétně studiem struktury, významu a užití slovních spojení. Základním principem frazeologie je přístup k jazyku ne jako k systému postavenému pouze na kombinaci gramatických pravidel a slovní zásoby, ale jako k systému, v němž je slovní zásoba úzce a neoddělitelně propojena s kontextem (přesněji s kotextem). Ustálená slovní spojení tudíž představují nesmírně důležitou součást jazyka, a to jak z pohledu osvojení si jazyka, tak z pohledu produkce jazyka. Ustálená slovní spojení se obecně definují jako spojení dvou či více slov, která jsou určitou měrou lexikálně a / nebo sémanticky fixní. Na základě mnoha výzkumů a studií v této oblasti bylo zjištěno a dokázáno, že se velká většina slovní zásoby rodilých mluvčích skládá právě z ustálených slovních spojení a díky studiím v oblasti korpusové lingvistiky bylo také prokázáno, že většinu přirozeného jazyka tvoří opakující se vzorce a spojení, která jsou ve velké míře ustálená. Následně je zde stručně nastíněna historie, a to včetně důvodů, které vedly k rozpoznání důležitosti

frazeologie a frazeologických jednotek jak z pohledu rodilých, tak i nerodilých mluvčích. Frazeologie jakožto lingvistická disciplína vzniká v 50. letech 20. století. Za jejím vznikem stojí strukturní lingvistika (především J. R. Firth), jež vidí jazyk jako strukturu (stavbu, celistvost) výrazových prostředků, ustálených a účelných...¹⁰ a zaměřuje se na úzké propojení slovní zásoby a kontextu. V 60. letech 20. století se frazeologie a s ní spojené koncepty opět odsouvají do pozadí díky Chomského generativní gramatice, nicméně v 80. letech 20. století se frazeologie opět posouvá do popředí zájmu, a to hlavně díky vzniku tří nových lingvistických disciplín, tj. kognitivní lingvistiky, konstrukční gramatiky a korpusové lingvistiky. Základní principy a jednotky každé z těchto disciplín, stejně jako jejich spojitost s frazeologií, jsou v podkapitole stručně nastíněny.

Podkapitola 2.1.2. se zabývá dvěma hlavními přístupy k frazeologii jako takové, konkrétně tradičním, frazeologickým (tzn. funkčním) přístupem a na frekvenci založeným distribučním (tzn. kvantitativním) přístupem. Tradiční přístup k frazeologii se zabývá převážně fixnějšími ustálenými slovními spojeními, zatímco distribuční přístup zaujal mnohem širší perspektivu a zahrnuje i typy kombinací, které nejsou v tradičním přístupu posuzovány jako součást frazeologie. Tradiční přístup k frazeologii má kořeny v Sovětském svazu a dalších zemích východní Evropy a jedním z následovníků této tradice je A. P. Cowie. Frazeologie se v tomto přístupu chápe jako kontinuum s fixními ustálenými slovními spojeními na jedné straně a transparentními, variabilními ustálenými slovními spojeními na straně druhé. Jádrem tradičního přístupu tvoří idiomy a cílem tohoto přístupu je definovat kritéria a následně odlišit různé typy ustálených slovních spojení a najít hranici mezi ustálenými slovními spojeními a volnými spojeními, která nespádají pod rámec frazeologie. Hlavním kreditem této tradice je nepochybně zavedení terminologie a kritérií pro kategorizaci a odlišení různých typů ustálených slovních spojení, což umožnilo jejich následné zkoumání a analýzu. Druhý přístup k frazeologii, tzn. distribuční přístup, je úzce spojen s korpusovou lingvistikou a hlavně s frekvencí souvškytu. Základní myšlenkou této tradice je, že ustálená slovní spojení (nehledě na jejich povahu) jsou mnohem důležitější než jednotlivá slova (a mají před nimi přednost). Hlavním představitelem této tradice je J. Sinclair, který nehodnotí ustálená slovní spojení na základě čistě lingvistických

¹⁰ Vachek, 154.

kritérií, ale zaměřuje se právě na frekvence souvýskytu slov na základě korpusové lingvistiky. Díky tomuto přístupu se do popředí zájmu dostávají kolokace, koligace a jiná ustálená slovní spojení (které v tradičním přístupu patří mimo rámec frazeologie), která se díky korpusové lingvistice ukázala být mnohem častějším jevem v jazyce než fixnější ustálená slovní spojení (tvořící jádro tradičního přístupu). Všechny tyto typy ustálených slovních spojení jsou součástí Sinclairova „idiom principle“, na kterém je založen jazyk a slouží jako protipól jeho „open-choice principle“.

Kapitola 3. věnuje pozornost ustáleným slovním spojení jako takovým a zabývá se převážně popisem jednotlivých typů těchto spojení se zaměřením na kolokace. Hlavním problémem v této oblasti je nejen to, že existuje hodně různých typů ustálených slovních spojení, ale hlavně, že neexistuje jednotný přístup k tomuto jevu. Jak již bylo zmíněno, existují dva hlavní přístupy k frazeologii, a tudíž existují také různé typologie ustálených slovních spojení. Tyto typologie jsou úzce spojeny s vybraným přístupem k frazeologii. Jinými slovy, výběr a pořadí kritérií použitých ke kategorizaci ustálených slovních spojení a k odlišení různých typů těchto spojení je úzce spjato s vybraným přístupem k frazeologii obecně. Podkategorie 3.1. pro příklad uvádí dva seznamy kritérií používaných pro kategorizaci a následné odlišení různých typů ustálených slovních spojení. Granger & Paquot uvádí pět základních kritérií pro kategorizaci a odlišení různých typů těchto spojení jsou interní struktura, rozsah, sémantika, syntaktická flexibilita a diskurzivní funkce. Pro porovnání, Griesův seznam obsahuje kritérií šest a přestože se částečně shoduje s kritérii navrženými Granger & Paquot, nebere v potaz diskurzivní funkci spojení a naopak přidává frekvenci a přípustnou vzdálenost mezi prvky do seznamu kritérií. Podkategorie 3.2. popisuje různé přístupy týkající se kategorizace a rozlišení různých typů ustálených slovních spojení. Pro příklad jsou uvedeny tři významné typologie, které měly důležitý vliv na ustálená slovní spojení z pohledu jejich kategorizace a rozlišení různých typů těchto spojení. První takovou typologií, důležitou z hlediska anglické lexikologie a lexikografie, je typologie Cowieho, jejíž hlavním přínosem je primární rozdělení ustálených slovních spojení z hlediska sémantického + syntaktického a pragmatického. Tímto odlišuje tzv. „composites“ obsahující kolokace a idiomy a tzv. „formulae“, která mají funkci pragmatickou. Další důležitou typologií je typologie Mel'čuka, která je víceméně shodná s Cowieho typologií, nicméně Mel'čuk používá jinou terminologii. Důležitým prvkem Mel'čukovy typologie je popis kolokací na základě lexikálních funkcí, konkrétně pokus o popsání lexikálních preferencí (tzn. proč jsou

některé kolokace přirozené a přípustné, zatímco ostatní ne) pomocí lexikálních funkcí. Třetí důležitou typologií je Burgerova typologie. Oproti předchozím dvěma typologiím se tato primárně zaměřuje na ustálená slovní spojení z hlediska diskurzu a diskurzivních funkcí. Burger rozděluje ustálená slovní spojení podle tří hlavních kategorií, tzn. referenční, strukturní a komunikativní ustálená spojení. Podkategorie 3.3. popisuje pokus Granger & Paquot o sjednocení dvou přístupů k frazeologii (a tudíž i o sjednocení typologií týkající se ustálených slovních spojení). Granger & Paquot navrhuje v první řadě sjednocení (a tím pádem zjednodušení) různých terminologií a navrhuje vytvořit dvě různé terminologie, jednu pro automatickou extrakci ustálených slovních spojení (tzv. distribuční kategorie) a druhou pro lingvistické analýzy těchto spojení (tzv. frazeologické spektrum).

Kapitola 4. se zaměřuje na kolokace, konkrétně na různé klasifikace a definice tohoto pojmu (opět spojené s vybraným přístupem k frazeologii), na jejich rozlišení od ostatních ustálených slovních spojení a také představuje definici kolokace relevantní pro analytickou část této práce. Jak již bylo zmíněno, existují dva hlavní přístupy k frazeologii a následně i k ustáleným slovním spojení včetně kolokací. Podkapitola 4.1. shrnuje hlavní přístupy ke kolokacím (a z nich vyplývajících klasifikace a definice). Podle distribučního přístupu je kolokace definována na základě souvýskytu slov v určité vzdálenosti (bez ohledu na syntaktický vztah mezi nimi), a hlavní rozdíl je mezi souvýskytami, které jsou frekventované (tzn. častější, než kdyby se tato slova kombinovala náhodně) a souvýskytami nefrekventovanými. Sinclair definuje kolokaci v zásadě jako lexikální souvýskyt dvou či více slov ve vzájemné blízkosti a pro lepší pochopení terminologie také uvádí definici idiomu. Tradiční, frazeologický přístup definuje kolokaci jako kombinaci slov, která je do určité míry fixní (ale ne úplně). Cowie klasifikuje kolokaci jako jeden z typů tzv. „composites“ (mající primárně syntaktickou + sémantickou funkci) a kolokace je definována jako kombinace slov, z nichž alespoň jedno má přenesený / idiomatický význam a alespoň jedno je užito v doslovném významu a celá kombinace slov je transparentní. Dále jsou stručně uvedeny také další, nicméně méně významné, přístupy ke kolokacím. Podkapitola 4.2. popisuje dvě hlavní kritéria k popisu a rozlišení kolokací od dalších typů ustálených slovních spojení, konkrétně tzv. předpoklady a kontinua. Vzhledem k tomu, že se kolokace obecně definuje jako souvýskyt dvou nebo více slov, předpokladem je, že tato dvě slova musí být přístupná kombinacím a zároveň se musí vyskytovat ve stejném kontextu (lépe řečeno kotextu). Mezi kontinua se řadí

sémantická transparence, kolokační řada a frekvence souvýskytu. Finálně, podkapitola 4.3. se zabývá výhradně definicí a klasifikací kolokací relevantní pro tuto práci. Definice i klasifikace kolokací byly přímo převzaty z frazeologického spektra navrženého Granger & Paquot. Pro ilustraci a snadnější orientaci v terminologii jsou zde definovány jak lexální, tak i gramatické kolokace, a také je zde definována kategorie idiomů. Vzhledem k tomu, že v analytické části této studie je zkoumán pouze jeden určitý typ kolokací, tzn. kombinace anglických příslovcí končících na *-ly* a přídavných jmen, podkapitola 4.3.1. věnuje pozornost právě kategorii těchto anglických intenzifikačních příslovcí, tzv. amplifiers.

Pátá kapitola se zabývá dosavadním výzkumem kolokací v angličtině a shrnuje základní výsledky studií kolokací u studentů angličtiny. Tyto studie se shodují na faktu, že kolokace všeobecně představují problém pro studenty angličtiny a také, že je nezbytné, aby kolokace dostaly mnohem větší prostor a důležitost při výuce, tzn. v procesu osvojování si, jazyka. Podkapitola 5.1. detailně popisuje studii Sylvianne Granger, na které je založena analytická část této práce (viz výše), včetně předpokladů a metodologie a představuje výsledky této studie.

Kapitola 6. představuje seznámení s metodikou práce. Jak již bylo zmíněno, analytická část této práce je založena na testování anglických kolokací u českých studentů angličtiny na základě podobné studie Sylvianne Granger. Testování českých studentů angličtiny bylo založeno na dvou různých typech testů, tzv. „Cloze test“ a „Significance of Collocation test“. První typ testu byl založen na elicitaci, tzn. doplňování 20 vět minimálně jedním a maximálně deseti anglickými intenzifikačními příslovcí končícími na *-ly*. Druhý test byl přímo převzat ze studie Granger a tazatelé měli za úkol ke každému z 11 daných intenzifikačních příslovcí vybrat ze seznamu 15 přídavných jmen pouze ta, která se podle jejich názoru dala kombinovat s daným příslovcem. Výsledky z obou testů byly následně porovnány s výsledky z Britského národního korpusu, které nahrazovaly rodilé mluvčí v této studii a sloužily jako „ideální“ stav pro porovnání výsledků českých studentů angličtiny.

Sedmou a zároveň nejdůležitější kapitolou této práce je analýza dat, založená na výsledcích výše zmiňovaných testů a datech extrahovaných z Britského národního korpusu. Základní hypotézou analýzy je předpoklad, že znalosti a praktické použití anglických kolokací budou u českých studentů angličtiny srovnatelně horší než u rodilých mluvčích, konkrétně že studenti angličtiny budou používat méně kolokací než rodilí mluvčí, že konkrétní kolokace se budou lišit z pohledu preferencí u obou skupin,

a také že se u českých studentů angličtiny vyskytne množství chyb, jak co se týče znalostí, tak i praktického použití anglických kolokací. Analytická část je rozdělena na čtyři hlavní podkapitoly. Podkapitola 7.1. popisuje anglické intenzifikační příslovce končící na *-ly* z obecného hlediska, tzn. uvádí 15 nejčastěji používaných anglických příslovcí tohoto typu v Britském národním korpusu. Podkapitoly 7.2. a 7.3. detailně popisují a vyhodnocují výsledky testů použitých pro analýzu. V podkapitole 7.2. jsou prezentovány výsledky „Cloze testu“. Oproti původnímu předpokladu, že studenti angličtiny budou používat méně kolokací než rodilí mluvčí, test ukázal pouze malé rozdíly v počtu příslovcí navrhovaných českými studenty a výsledky z Britského národního korpusu. Je třeba nicméně brát v úvahu fakt, že pro analýzu bylo použito jen 5 nejčastějších příslovcí z *BNC* ke každému přídavnému jménu. Jinými slovy, pokud by byly brány v potaz všechny možné kombinace v *BNC*, původní předpoklad by se potvrdil. Na základě prvních analýz studie potvrdila markantní rozdíly z pohledu výběru a preferencí jednotlivých příslovcí u obou skupin. Čeští studenti angličtiny používali v zásadě stejná příslovce bez ohledu na přídavné jméno v dané kombinaci, tudíž nejen, že jejich cit pro salienci (tzn. aktivovanost) byl v porovnání s rodilými mluvčími srovnatelně slabší, ale také jejich kreativní schopnosti se ukázaly být značně limitovány ve srovnání s druhou skupinou. Ve výběru a preferencí jednotlivých příslovcí se také projevil velice silný vliv mateřského jazyka, tzn. češtiny. Nutno dodat, že vliv češtiny je v tomto případě převážně negativní. Detailní rozbor rozdílu v užívání kolokací u obou skupin naznačily dvě tendence. Zaprvé, čeští studenti angličtiny prokázali relativně dobrou znalost nejfrekventovanějších, běžně používaných (a v zásadě prototypických) anglických kolokací. Na druhou stranu, jejich znalosti a schopnosti používat anglické kolokace se z obecného hlediska ukázaly jako poměrně slabé. Hlavním faktorem je zde pravděpodobně vliv mateřského jazyka, který je zodpovědný za velké množství chyb v používání anglických kolokací a stejně tak za poměrně častý výskyt v angličtině neexistujících kolokací navrhovaných českými studenty angličtiny. Podkapitola 7.2. shrnuje a rozebírá výsledky „Significance of collocation testu“. Přestože byl tento test v porovnání s předchozím testem jednodušší, tzn. nebyl založen na elicitaci, ale pouze na výběru přídavných jmen z daného seznamu (tzn. na salienci), výsledky českých studentů angličtiny v tomto typu testu byly taktéž nedostačující. Prvním poznatkem bylo, že počet kombinací navrhovaných českými studenty ztelně převyšoval počet kombinací extrahovaných z *BNC*. Jinými slovy, čeští studenti angličtiny považovali za

přijatelných téměř o sto kombinací více, než bylo ve skutečnosti nalezeno v korpusu. Tento výsledek opět ukazuje na fakt, že salience / aktivovanost je u nerodilých mluvčích výrazně slabší, ale také částečně chybná. Opět je zde patrný vliv mateřského jazyka. Tazatelé měli v testu za úkol nejen vybrat všechny vhodné, přijatelné, kombinace, ale také zakroužkovat pro každé příslovce pouze jedno přídatné jméno, které vykazovalo s daným příslovcem největší míru salience, tzn., které se k danému příslovci vyskytovalo podle jejich názoru nejčastěji ze všech, tudíž s ním bylo nejtěsněji spjato. V tomto případě byly výsledky českých studentů angličtiny o poznání lepší (přestože i zde se objevily případy kolokací, které v angličtině neexistují). Stejně jako v případě prvního testu, výsledky „Significance of collocation testu“ naznačily dvě tendence. Čeští studenti angličtiny opět prokázali dobrou znalost frekventovaných, běžně se vyskytujících (a v zásadě prototypických) kolokací s vysokou mírou salience. Na druhé straně, méně frekventované kolokace se opět ukázaly jako velice problematické pro české studenty angličtiny a opět se projevil silný vliv mateřského jazyka na základě poměrně velkého počtu navrhovaných kombinací, které se ale v Britském národním korpusu nevyskytují. Výsledky obou testů v zásadě potvrdily původní předpoklady, čeští studenti angličtiny neprojevili nijak zásadně uspokojující znalosti anglických kolokací (samozřejmě s výjimkou těch nejběžnějších) - na základě analýzy těchto výsledků se zdá, že čeští studenti angličtiny používají anglické intenzifikační příslovce jako jednotlivá slova, která se mohou kombinovat s teoreticky jakýmkoliv přídatným jménem, než jako součást ustálených slovních spojení. Otázkou ovšem zůstává, do jaké míry jsou výsledky analýzy opravdu nedostačující. Na jedné straně je možno argumentovat, že znalost nejfrekventovanějších, běžně používaných, anglických kolokací je pro české studenty dostačující, že není potřeba znát všechny možné kolokace – což je do jisté míry pravda. Na druhou stranu, nedostatek znalostí o ostatních možných, ale hlavně nemožných, anglických kombinacích může vést např. k nedorozumění v interakci s rodilými mluvčími. Z obecného hlediska je míra chybovosti u českých studentů angličtiny tak vysoká, že je nemožné ji ignorovat a tudíž se výsledky testování kolokací u českých studentů angličtiny dají jen těžko označit za uspokojivé, natož dobré. Podkapitola 7.4. stručně nastiňuje pedagogické implikace.

Appendix

Cloze test (CT 004)

Age: 26

Sex: Male - Female

Your first language: Czech

How long have you been learning English? 15 years

How much time have you spent in an English-speaking country? 3 months

Instructions: Complete the following sentences with at least one amplifier (max. 5-10), i.e. an adverb ending in *-ly* expressing a **high degree**.

1. *The news was -----ly good for me.*

1.	Really	2	Incredibly	3	
4.		5		6	
7.		8		9	
10					

2. *Those years were -----ly great.*

1.	Really	2	Absolutely	3	
4.		5		6	
7.		8		9	
10					

3. *His approach is -----ly different from mine.*

1.	Totally	2	Completely	3	
4.		5		6	
7.		8		9	
10					

4. These issues are -----ly important for researchers.

1.	Highly	2	Extremely	3	
4.		5		6	
7.		8		9	
10					

5. The situation is -----ly bad for Britain's industry.

1.	Particularly	2	Especially	3	
4.		5		6	
7.		8		9	
10					

6. This task is -----ly difficult for many people.

1.	Extremely	2	Really	3	
4.		5		6	
7.		8		9	
10					

7. This is -----ly true for the young.

1.	Particularly	2	Especially	3	
4.		5		6	
7.		8		9	
10					

8. It is -----ly necessary to consider the arguments used in the public debates.

1.	Extremely	2		3	
4.		5		6	
7.		8		9	
10					

9. Some topics can be -----ly hard to understand.

1.	Really	2	Awfully	3	
4.		5		6	
7.		8		9	
10					

10. That solution was -----ly simple.

1.	Incredibly	2	Unbelievably	3	
4.		5		6	
7.		8		9	
10					

11. This feeling is -----ly natural.

1.	Absolutely	2	Completely	3	Totally
4.		5		6	
7.		8		9	
10					

12. At the wedding she seemed -----ly happy.

1.	Really	2	Incredibly	3	
4.		5		6	
7.		8		9	
10					

13. The situation has become -----ly serious.

1.	Really	2	Extremely	3	
4.		5		6	
7.		8		9	
10					

14. These changes have had a -----ly significant impact on the local economy.

1.	Highly	2		3	
4.		5		6	
7.		8		9	
10					

15. It made him -----ly aware of the danger.

1.	Fully	2		3	
4.		5		6	
7.		8		9	
10					

16. The President personally remained -----ly popular.

1.	Highly	2	Extremely	3	Really
4.		5		6	
7.		8		9	
10					

17. Last winter was -----ly cold.

1.	Incredibly	2	Really	3	
4.		5		6	
7.		8		9	
10					

18. This information will be -----ly useful for him.

1.	Extremely	2	Really	3	
4.		5		6	
7.		8		9	
10					

19. That mission seemed -----ly impossible.

1.	Utterly	2		3	
4.		5		6	
7.		8		9	
10					

20. At the age of 16 he became -----ly ill.

1.	Seriously	2		3	
4.		5		6	
7.		8		9	
10					

21. In practice, no witnesses are -----ly reliable.

1.	Completely	2		3	
4.		5		6	
7.		8		9	
10					

22. I felt sad, depressed, -----ly miserable.

1.	Absolutely	2	Totally	3	Completely
4.		5		6	
7.		8		9	
10					

23. I'm -----ly sorry.

1.	Really	2		3	
4.		5		6	
7.		8		9	
10					

Significance of collocation test

Sex: Male - Female

Age: 23

How long have you been learning English? about 8 years

Have you spent some time in an English-speaking country? How long? a month in Scotland

Instruction: From a list of 15 adjectives in each case, choose the acceptable collocates of 11 amplifiers. Underline all the adjectives which in your opinion collocate with the amplifier and circle the adjective which in your opinion is more frequently associated with the amplifier than all the others. Thank you!

highly	<u>significant</u>	<u>reliable</u>	ill	<u>different</u>	<u>essential</u>	aware	<u>miserable</u>	available
	difficult	happy	clear	ignorant	<u>impossible</u>	cold	<u>important</u>	
seriously	significant	reliable	<u>ill</u>	different	essential	aware	miserable	available
	difficult	happy	clear	ignorant	impossible	cold	<u>important</u>	
readily	<u>significant</u>	<u>reliable</u>	ill	<u>different</u>	essential	aware	<u>miserable</u>	<u>available</u>
	difficult	happy	clear	ignorant	impossible	cold	important	
blissfully	significant	reliable	ill	different	essential	aware	miserable	available?
	difficult	happy	clear	ignorant	impossible	cold	<u>important</u>	
vitality	<u>significant</u>	<u>reliable</u>	ill	<u>different</u>	essential	aware	<u>miserable</u>	available
	difficult	happy	clear	ignorant	impossible	cold	<u>important</u>	
fully	significant	<u>reliable</u>	ill	<u>different</u>	essential	<u>aware</u>	miserable	available
	difficult	happy	clear	ignorant	impossible	cold	important	
perfectly	significant	reliable	ill	<u>different</u>	essential	<u>aware</u>	miserable	available
	difficult	<u>happy</u>	<u>clear</u>	ignorant	<u>impossible</u>	cold	important	
heavily	significant	reliable	<u>ill</u>	different	essential	aware	miserable	available
	difficult	happy	clear	<u>ignorant</u>	impossible	cold	important	
bitterly	significant	reliable	ill	different	essential	aware	<u>miserable</u>	available
	difficult	happy	clear	<u>ignorant</u>	impossible	<u>cold</u>	important	
absolutely	<u>significant</u>	reliable	ill	<u>different</u>	essential	aware	miserable	available
	difficult	happy	<u>clear</u>	ignorant	<u>impossible</u>	cold	important	
utterly	significant	reliable	ill	different	<u>essential</u>	aware	miserable	available
	<u>difficult</u>	happy	clear	<u>ignorant</u>	<u>impossible</u>	cold	important	

Significance of collocation test

Sex: Male - Female

Age: 20

How long have you been learning English? 10 years

Have you spent some time in an English-speaking country? How long? 1 year in the US

Instruction: From a list of 15 adjectives in each case, choose the acceptable collocates of 11 amplifiers. Underline all the adjectives which in your opinion collocate with the amplifier and circle the adjective which in your opinion is more frequently associated with the amplifier than all the others. Thank you!

highly	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill <u>clear</u>	<u>different</u> ignorant	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
seriously	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	<u>ill</u> clear	<u>different</u> ignorant	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
readily	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill clear	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
blissfully	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill clear	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
vitaly	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill clear	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
fully	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill clear	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
perfectly	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill <u>clear</u>	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
heavily	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill clear	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
bitterly	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill clear	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
absolutely	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill clear	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>
utterly	<u>significant</u> <u>difficult</u>	<u>reliable</u> happy	ill clear	<u>different</u> <u>ignorant</u>	<u>essential</u> <u>impossible</u>	<u>aware</u> <u>cold</u>	<u>miserable</u> <u>important</u>	<u>available</u>

Significance of collocation test

Sex: Male - Female

Age: 20

How long have you been learning English? 14

Have you spent some time in an English-speaking country? How long? no

Instruction: From a list of 15 adjectives in each case, choose the acceptable collocates of 11 amplifiers. Underline all the adjectives which in your opinion collocate with the amplifier and circle the adjective which in your opinion is more frequently associated with the amplifier than all the others. Thank you!

highly	significant difficult	<u>reliable</u> happy	ill clear	different ignorant	essential impossible	aware cold	miserable important	available
seriously	significant difficult	reliable happy	<u>ill</u> clear	different ignorant	essential impossible	aware cold	miserable important	available
readily	significant difficult	reliable happy	ill clear	different ignorant	essential impossible	aware cold	miserable important	<u>available</u>
blissfully	significant difficult	reliable <u>happy</u>	ill clear	different <u>ignorant</u>	essential impossible	aware cold	miserable important	available
vitaly	significant difficult	reliable happy	ill clear	different ignorant	essential impossible	aware cold	miserable <u>important</u>	available
fully	significant difficult	reliable happy	ill clear	different ignorant	essential impossible	<u>aware</u> cold	miserable important	<u>available</u>
perfectly	significant difficult	<u>reliable</u> <u>happy</u>	ill <u>clear</u>	different ignorant	essential <u>impossible</u>	aware cold	miserable important	<u>available</u>
heavily	<u>significant</u> difficult	reliable happy	ill clear	different ignorant	essential impossible	aware cold	miserable important	available
bitterly	significant difficult	reliable <u>happy</u>	ill clear	different ignorant	essential impossible	aware <u>cold</u>	miserable important	available
absolutely	significant difficult	<u>reliable</u> <u>happy</u>	ill <u>clear</u>	different ignorant	essential impossible	aware cold	<u>miserable</u> important	available
utterly	significant difficult	<u>reliable</u> happy	ill clear	different ignorant	essential impossible	aware cold	<u>miserable</u> important	available

Significance of collocation test

Sex: Male - Female

Age: 21

How long have you been learning English? 12

Have you spent some time in an English-speaking country? How long? 1 year

Instruction: From a list of 15 adjectives in each case, choose the acceptable collocates of 11 amplifiers. Underline all the adjectives which in your opinion collocate with the amplifier and circle the adjective which in your opinion is more frequently associated with the amplifier than all the others. Thank you!

highly	<u>significant</u> difficult	<u>reliable</u> happy	ill clear	different ignorant	essential <u>impossible</u>	<u>aware</u> cold	miserable <u>important</u>	available
seriously	significant difficult	reliable happy	<u>ill</u> clear	different ignorant	essential impossible	aware cold	miserable important	available
readily	significant difficult	reliable happy	ill clear	different ignorant	essential impossible	aware cold	miserable important	<u>available</u>
blissfully	significant difficult	<u>reliable</u> happy	ill <u>clear</u>	different <u>ignorant</u>	essential impossible	<u>aware</u> cold	miserable important	available
vitality	<u>significant</u> difficult	reliable happy	ill clear	different ignorant	essential impossible	aware cold	miserable important	available
fully	significant difficult	reliable happy	ill clear	different ignorant	essential impossible	<u>aware</u> cold	miserable important	<u>available</u>
perfectly	significant difficult	<u>reliable</u> happy	ill <u>clear</u>	different ignorant	essential impossible	<u>aware</u> cold	miserable important	available
heavily	significant difficult	<u>reliable</u> happy	ill clear	different ignorant	<u>essential</u> impossible	aware cold	miserable important	available
bitterly	significant difficult	reliable happy	ill clear	different ignorant	essential impossible	aware <u>cold</u>	miserable important	available
absolutely	significant difficult	reliable happy	ill <u>clear</u>	different ignorant	<u>essential</u> <u>impossible</u>	aware cold	miserable <u>important</u>	available
utterly	significant difficult	reliable happy	ill clear	different ignorant	<u>essential</u> <u>impossible</u>	aware cold	miserable important	available