Inhaltsverzeichnis

Figures and Tables 11

Acknowledgements 15

1 Three sides of the same coin? – Collocations, Creativity, Constructions 17

2 Collocations as Constructions 25
2.1 Context-Oriented Approaches 27
2.2 Significance-Oriented Approaches 36
2.3 A Construction-Oriented Approach 48

3 Collocations and Creativity 55
3.1 Creative Variation of Collocations 56
3.2 Creativity and Cognition 64

4 Creating Linguistic Creativity 69
4.1 Nativist Approaches 71
4.1.1 Interpretive Semantics 73
4.1.2 Conceptual Semantics and Parallel Architecture 75
4.2 Constructionist Approaches 78
4.2.1 Emergentist Approaches 79
4.2.2 Social-pragmatic Learning 87
4.3 Phraseology and Language Acquisition 91
4.3.1 Construction Grammar 92
4.3.2 A Stage Model 100
4.3.3 Complex Adaptive Systems 104
4.4 The DMCDC-Model: A usage-based model of collocations 106
4.5 Summary and Implications 111

5 Measuring Collocations – Methodological Considerations 115
5.1 Online Production Tasks – Corpus Data and Statistical Association Measures for Collocations 118
5.1.1 Traditional Association Measures 120

5.1.2 Corpus Data in Cognitive Linguistic Research ... 125
5.2 Offline Perception Tasks – Experimental Data in Usage-based and Constructionist Studies ... 128
5.3 Methodological Limitations and Shortcomings ... 130
5.3.1 Corpus Data ... 130
5.3.2 Judgement Tasks ... 133
5.4 Methodology of this Study ... 134
5.4.1 Instruments ... 135
5.4.2 Participants and Procedure ... 141

6 CollMatch ... 145
6.1 Native Speakers – Adult ... 148
6.2 Native Speakers – Children ... 151
6.3 Native Speakers – Patterns ... 153
6.3.1 Pattern 1: Gradual Acceptance ... 153
6.3.2 Pattern 2: Peaked Acceptance ... 157
6.3.3 Pattern 3: Steady Acceptance ... 159
6.3.4 Pattern 4: Receding Positive Evaluation ... 161
6.3.5 Distractors ... 164
6.3.6 Summary ... 166
6.4 Non-Native Speakers ... 172
6.4.1 Pattern 1: Gradual Acceptance ... 175
6.4.2 Pattern 2: Peaked and Dented Acceptance ... 177
6.4.3 Pattern 3: Steady Acceptance ... 179
6.4.4 Pattern 4: Receding Positive Evaluation ... 180
6.4.5 Distractors ... 183
6.4.6 Summary ... 184
6.5 Effects of Schooling ... 188
6.6 Summary and Implications ... 193

7 CollJudge ... 197
7.1 Native Speakers ... 201
7.1.1 Pattern 1: Preference of Established Variants ... 201
7.1.2 Pattern 2: Overall Acceptance ... 207
7.1.3 Pattern 3: Contextual Acceptance ... 211
7.1.4 Other Patterns ... 213
7.1.5 Summary ... 215
7.2 Non-native speakers ... 216
7.2.1 Pattern 1: Preference of Established Variants ... 217
7.2.2 Pattern 2: Overall Acceptance ... 221

7.2.3 Pattern 3: Contextual Acceptance ... 223
7.2.4 Pattern 4: Contextual Influence ... 228
7.2.5 Other Patterns ... 231
7.2.6 Summary ... 234
7.3 Comparing Corpus Data and Evaluations from Judgement Tasks ... 236
7.4 Summary and Implications ... 242

8 Main Results and Implications ... 247
8.1 Main Results of this Study ... 248
8.2 Limitations and Further Research ... 253
8.3 Implications for a Usage-based Approach Towards language ... 255
8.3.1 First Language Acquisition ... 256
8.3.2 Second Language Acquisition and Learning ... 259

Appendices ... 265
Appendix I: Questionnaire ... 266
Appendix II: CollMatch (Acceptance Scores) ... 273
Appendix III: REC items vs. alternate combinations ... 277
Appendix IV: Raw Frequency Rankings and Association Measures for CollJudge ... 280
Appendix V: CollJudge (z-transformed acceptance scores) ... 281

References ... 283
Dictionaries ... 306
Corpora ... 306

Figures and Tables

Box 2.1: Entry for the lemma *do* from *Thousand-Word English* (Palmer / Hornby 1937: 43) 38
Box 2.2: Lexical functions of phraseological phenomena according to Mel'čuk (1995: 186) 43
Box 5.1: Test item from CollJudge – creative / simple variant of *cook the tea / meal* 138
Box 5.2: Code and example from the non-linguistic distractor task (Appendix I) 141
Box 6.1: Items of *Gradual Acceptance* (L1) 154
Box 6.4: Items with an overall tendency of receding positive evaluation (L1) 162
Box 6.5: CollMatch's pseudo-collocations according pattern (L1) 165
Box 6.7: Items of *Gradual Acceptance* (L2) 175
Box 6.8: Items of *Peaked Acceptance* and *Dented Acceptance* (L2) 178
Box 6.10: Items with an overall tendency of receding positive evaluation (L2) 181
Box 6.11: CollMatch's pseudo-collocations according pattern (L2) 183
Box 6.13: Items of *Academic Rejection* (L2) 187

Figure 1.1: Scope of this Study 24
Figure 2.1: Mel'čuk's (1989, 1995) process of text production compared to de Saussure's (1916 / 1967: 76–79) linguistic sign 42
Figure 2.2: Typology of word combinations (based on Hausmann 1984 and Bartsch 2004: 38) 44
Figure 2.3: Form and function levels of collocational construction, taking *scornful tone* as an example 51
Figure 4.1: The productivity cline according to Bra∂dal (2008: 172) 99
Figure 4.2: A stage model for the acquisition of formulaic language (Wray / Perkins 2000) 101
Figure 4.3: A dynamic model for the cognitive development of collocations (DMCDC) 106

Graph 6.1: Schematic overview of CollMatch's collocations and pseudo-collocations evaluated by adult native speakers of English (see Appendix II for a detailed overview) 149
Graph 6.2: Example for *Gradual Acceptance* (GA) – L1 acceptability rating for *drop hints* 154
Graph 6.3: Example for *Peaked Acceptance* (PA) – L1 acceptability rating for *meet a need* 157
Graph 6.4: Example for *Steady Acceptance* (StA) – L1 acceptability rating for *pull a face* 159
Graph 6.5: Example for Academic Acceptance (AcA) – L1 acceptability rating for *adopt an approach* 169
Graph 6.6: Example for *Dented Acceptance* (DA) – L2 acceptability rating for *make a move* 178
Graph 7.1: Example for *Preference of Established Variants* (PREF) – L1 acceptability rating for *pull a face* 202
Graph 7.2: Example for *Overall Acceptance* (OA) – L1 acceptability rating for *lend support/advice* 207
Graph 7.3: Example for *Contextual Acceptance* (CONTEXT) – L1 acceptability rating for *commit a crime/mistake* 211

Table 2.1: Lexical and grammatical relations (based on Halliday 1966: 152–153) 30
Table 2.2: Corpus-based association measures for "Humpty Dumpty's collocations" 34
Table 2.3: Categorisation of lexical word combinations according to Cowie (1983: xii-xiii) with examples from Howarth (1996:15–16) 39
Table 2.4: Categorisation of *open collocations* (Cowie/Howarth 1995: 83) ... 41
Table 2.5: List of noun collocates for *scornful* according to the BNC 46
Table 5.1: Typology of linguistic methodologies (partly based on Siyanova/Schmitt 2008) 116
Table 5.2: Basic contingency table for observed and expected frequencies (based on Evert 2009: 1231) 119
Table 5.3: Ten most frequent verb collocates for the lemma *crime* according to MI, z-score, t-score and log-likelihood (BNC) 122
Table 5.4: Distribution of participants across languages and age groups ... 142
Table 6.1: Overview of group results from CollMatch based on Gyllstad (2007) 146
Table 6.2: Patterns identified in CollMatch based on each L1 group's acceptance scores 167

Table 6.3: Overview of items with the same acceptance pattern or a similar acceptance score in adult L1 and L2 evaluations 173
Table 6.4: Patterns identified in CollMatch based on each L2 group's acceptance scores 185
Table 6.5: Overview of group results from CollMatch in year 5 (Germany) and year 7 (Great Britain) 191
Table 6.6: Comparison of group results from CollMatch in year 5 (L2) and year 7 (L1) 192
Table 7.1: Items in CollJudge in their four different variants (original version of each item in bold print) 198
Table 7.2: CollMatch results according to the four test variants of CollJudge for adult L1 and L2 participants 200
Table 7.3: L1 speakers' qualitative evaluation of items with the pattern of Preference of Established Variants sorted according to age and variant .. 204
Table 7.4: L1 speakers' qualitative evaluation of items with the pattern of Overall Acceptance sorted according to age and variant 209
Table 7.5: L1 speakers' qualitative evaluation of items with the pattern of Contextual Acceptance sorted according to age and variant 212
Table 7.6: L1 speakers' qualitative evaluation of items with an unclear pattern sorted according to age and variant 214
Table 7.7: L2 speakers' qualitative evaluation of items with the pattern of Established Variants sorted according to age and variant 219
Table 7.8: L2 speakers' qualitative evaluation of items with the pattern of Overall Acceptance sorted according to age and variant 221
Table 7.9: L2 speakers' qualitative evaluation of items with the pattern of Contextual Acceptance sorted according to age and variant 225
Table 7.10: L2 speakers' qualitative evaluation of items with the pattern of Contextual Influence sorted according to age and variant 230
Table 7.11: L2 speakers' qualitative evaluation of items with an unclear pattern sorted according to age and variant 233
Table 7.12: Raw frequencies and association measures for CollJudge Items 237
Table 7.13: Contextual influence (Δp) on the evaluation of native and non-native speakers of English for established and creative variants within CollJudge 241

Acknowledgements

Very much like collocations, this book was part of a steady process of 'creativity', 'construction' and, of course, 'change'. Many people helped me to pass through the different (and sometimes difficult) stages applied research involves. There are too many to name them all, but I am grateful for their support and feedback. Some, however, deserve a very special mention:

First, I would like to thank Professor Dr. Thomas Herbst and Professor Dr. Thorsten Piske for their valuable and constructive suggestions during all stages of this project.

I also wish to express my sincere gratitude to Dr. Christina Schelletter for helping out when help was needed the most. Furthermore, heartfelt thanks to Dr. Susanne Dyka, Dr. Carolin Ostermann, Eva Scharf and Thomas Binder for their feedback, encouragement and friendship. This book would not exist without them. I am also grateful to my colleagues as well as to all students, teachers and parents on both sides of the channel who, quite literally, provided the foundations for this project.

Finally, I wish to thank my family for their patience, support and encouragement. They are my 'base' and inspiration.

1 Three sides of the same coin? – Collocations, Creativity, Constructions

> Putting together novel expressions is something that speakers do, not grammars. It is a problem-solving activity that requires a constructive effort on the part of a speaker and occurs when he puts linguistic convention to use in specific circumstances.
> (Langacker 1987: 85)

Speakers, as Langacker (1987: 85) points out, form one of the key elements of a language. Not only are they the force that brings a language to life but also the motor to shape conventions and create new expressions, phrases, or even grammatical constructions. Even aspects of language which are traditionally defined through their invariability or at least partial fixedness, such as idioms or collocations, are not immune to creative alternations and change, as sentences[1] (1) to (3) show.

(1)	BNC $_{\text{K51 1683}}$	Politicians seem to work on the assumption that the early bird catches the voter.
(2)	BNC $_{\text{CJA 2253}}$	[...] and Tabitha was captivated despite herself, watching the pretty man play and wondering how he would end it, how he could ever resolve the disagreement between the rush and the ebb [...]
(3)	BNC $_{\text{H8H 1277}}$	Now I have, and I'm telling you that if you marry him then you'll be committing the biggest mistake in your brief little life.

This conflict between fixedness and change forms the basis for this study. The following pages focus on the apparent tension between established, idiomatic items, and creative alternations, which ultimately cause a language to change

1 The examples at hand are taken from the British National Corpus (BNC) with the relevant elemets underlined for the purpose of this study. In terms of frequency, sentences (1) and (2) are the only instances of *the early bird catches the voter* and *pretty man* respectively, while phrases containing the lemma *mistake* in combination with the lemma *commit* (with a span of +/- 4) occur more frequently, with a total of eight acceptable hits within the corpus.

and evolve. Since idioms[2], like *the early bird catches the worm,* are usually considered to function more or less as one complex, rather invariant, unit of meaning, which is made up of several words[3] but expresses a unified concept, it might seem rather surprising to find creative alternations as in (1). However, collocations, like *pretty woman* or *commit a crime*[4], play an even more interesting role. Most definitions (> 2) would, in fact, agree that the essence of this phraseological phenomenon is a strong, partly inexplicable, bond, which seems to link all items within a collocation. In fact, these collocates were often argued to be so closely associated that the thought or perception of one collocate almost automatically seems to somehow activate the other(s)[5], like *commit* would trigger a word such as *crime,* while *pretty* is likely to elicit *woman.* Therefore, to talk about creative, "novel expressions" in the context of collocations seems to be counter-intuitive at first. Nevertheless, for most definitions a second decisive feature of collocations is a certain degree of flexibility within their components, as in *pretty woman*, with alternations like *pretty girl* or *pretty face,* or *commit* which, amongst others, can also be found with *offence* or *act.* In some cases, these collocational combinations can then even extend to rather unlikely or even novel, yet decodable, combinations like *pretty man* or *commit a mistake.*

2 For a more detailed discussion of idioms and their relationship to collocations compare Wulff (2010: 11–12); Howarth (1996: 47); Cruse (1986); and Cowie / Mackin (1983: viii-ix).

3 As the examples above show, the notion of *word* could, at times, be misleading, since it is not a clearly defined concept. Therefore, it might apply to a sequence of letters or phonemes, separated from other sequences by a space or a pause – such as *bear* and *bird* or *bears* or *birds* – but also as a sequence of these items which constitute one concept such as *teddy bear* or *the early bird catches the worm.* Hence, Matthews (1974) suggests using the term *lexeme* for a unit referring to "fundamental" concepts while keeping the term *word* for a lexeme's instantiations. According to this distinction, *teddy bear* could then also be regarded as a *multi-word lexeme.*

4 These examples have been repeatedly used to discuss collocations and collocative meaning (compare for example Leech 1974; 1998: 20; Palmer 1976: 95–97; Klotz 1998: 92–95).

5 Herbst (1996), for example, did show that native speakers of English tend to agree on similar solutions in a completion task, but also that, compared to non-native participants, native speakers give a smaller range of alternative collocates. Hoffmann and Lehmann (2000) furthermore compared native as well as non-native speakers' intuition against corpus data for low-frequency collocations from the BNC and found that most of the native speakers scored above chance and were able to predict most of the missing collocates in a fill-in-the-gap task. In an reproduction task, Conklin and Schmitt (2008) focused on cognitive processing of formulaic sequences; most of their native as well as non-native participants were able to process formulaic sequences faster in a self-paced reading task, which further supports the assumption that formulaic combinations like collocations are, to a certain extent, cognitively linked.

However, as the examples in (2) and (3) show, a *pretty man* is not necessarily the same as the linguistically much more frequent[6] *handsome man*. Furthermore, if one *commits a mistake*, this is very unlikely to be the same kind of act as in to *make a mistake*; and a *pretty man* tends to be associated with rather female features or behaviour, often used in a derogatory or objectified way, like for example the suitor of the emancipated and self-confident protagonist[7] in sentence (2). A mistake which is *committed*, on the other hand, is very likely to be a euphemistic phrasing to refer to a serious offence or, as in (3), a similarly life-changing, yet wrong, decision. In these cases *pretty* and *commit* seem to coerce their respective noun phrases (NP) into a reading which is much closer to their established collocational meaning than to the more common combinations with *handsome* or *make*. In past publications on collocations, these examples have been treated as separate phenomena. They have either been classified as some kind of deliberate, creative, literary form of language use (cf. Hausmann 1984 on *counter-creations*), more or less brushed aside as lexical idiosyncrasies or peripheral phenomena[8] (Chomsky 1965, Palmer 1976, Klotz 1998) or have not been mentioned at all. One notable exception is Mackin (1978), who advocates for a lexicographical description of phraseological language which not only focuses on the prototypical form of an idiom or collocation but also takes into account creative alternations, which he calls *nonce uses* (Mackin 1978: 163–164). Also, comprehensive studies[9] on collocations tend to focus on high frequent or

6 The frequencies for the respective lemma combinations from the BNC (span +/- 4) around the noun-collocate: *pretty+man* (2), *handsome+man* (151), *commit+mistake* (8), *make+mistake* (2002).

7 The book sentence (2) is taken from is the award-winning SciFi-novel *Take Back Plenty* (Greenland 1990: 81) which features an extraordinary, strong and in every aspect of her life independent woman as its main protagonist.

8 Approaches within the framework of generative grammar tend to avoid phraseological phenomena in general and collocations in particular. This is largely because in a generativistic tradition semantic phenomena and their distributional properties are regarded as a peripheral phenomenon. In his publication on *Aspects of the Theory of Syntax*, for example, Chomsky (1965) suggests two sets of rules; "strict subcategorisation rules" and "selectional rules" (Chomsky 1965: 95). For Chomsky, violation of selectional rules, like for example *commit a mistake* instead of *make a mistake*, would still lead to a decodable syntactic structure and can thus be regarded as a less central aspect within a linguistic system. Rögnvaldsson (1993), on the other hand, argues that even though they are usually not explicitly accounted for in universalistic publications, later versions of Chomsky's conception of language especially, such as *The Minimalist Program* (Chomsky 1995), can provide a suitable framework to explain collocational combinations.

9 As for example in Bahns (1996) and Steinbügl (2005) with a focus on lexicography; Lehr (1996), Nesselhauf (2004) and Bartsch (2004) with a focus on corpus linguistic methods or Howarth (1996), Gyllstad (2007) and Jehle (2007) with a focus on language attainment and testing.

highly associated collocational pairs. This leaves more creative versions out of the picture, since, if viewed individually, for example in a large corpus like the British National Corpus (BNC[10]), they are often a low-frequency phenomenon. Even in association measures like *mutual information* (MI), these combinations lose out against rare collocates, like *cottage-residence* or *hara-kiri*. At the same time, these non-standard alternations of collocations are particularly interesting, because they are infrequent yet not incomprehensible, which shows that, to a certain extent, creativity and change need a base of established linguistic structures to be interpreted against. The most basic level which could be assumed would, of course, be simple syntactic rules in a traditional grammar-lexicon model (Chomsky 1965). Here, it would be argued that the examples of *pretty man* and *commit a mistake* are but a mere combination of two lexical items which are formed ad hoc on the basis of established syntactical structures, such as adjective plus noun, [Adj+N][11], or verb phrase plus noun phrase in object position, [VP+NP]. Still, this does not explain why these words are then interpreted against the background of a related, more established, actual collocation. Therefore, these instances beg the question, whether the interpretation of these creative word combinations might not indeed be cognitively supported by more common, entrenched collocational pairs. This connection, however, would imply that the reason different collocations operate on a gradient spectrum of fixedness is not just a linguistic fact or even coincidence, but that the degree of variability might depend on other factors, like the frequency of input or cognitive entrenchment[12].

Moreover, as our first interpretation of combinations like *pretty man* or *commit a mistake* demonstrated, these creative alternations might support approaches which suggest that not only traditional lexical items, such as words or compounds, but also more abstract constructions, such as [*pretty*+N] or [*commit*+NP], could have their own level of meaning. In the last decades, several cognitive and constructionist[13] approaches developed which have explicitly or implicitly committed themselves to this idea and thus regard language as a network of elements consisting of a formal as well as a functional side. Further-

10 Data cited with "BNC" has been extracted from the *British National Corpus*, distributed by Oxford University Computing Services on behalf of the BNC Consortium. All rights in the texts cited are reserved.

11 In this study, square brackets will be used to indicate that the respective sequence could be regarded as the formal representation of a construction. (cf. Stefanowitsch / Gries 2003)

12 *Entrenchment* in this chapter is used independently from any approach or school of thought. The discussion of *entrenchment* vs. *pre-emption* will later be part of chapter 4.

13 See Fischer / Stefanowitsch (2006), Ziem / Lasch (2013) or Hoffmann / Trousdale (2013) for a detailed overview of constructionist approaches.

more, they also see a speaker as the user as well as a source of linguistic innovation, and thus no longer distinguish between more or less normative language competence and a speaker's actual performance. Therefore, they instead see both the language and its speakers as inseparable parts of a dynamic system[14]. One prominent branch is usage-based theories, which have already been able to show specific effects of linguistic input and frequency on language attainment (Bybee 2010; Ellis 2006; Tomasello 2005; Bybee / Hopper 2001). Another school of thought, *construction grammar*, developed in recent decades and focuses explicitly on language as a system of form-function pairings, so-called *constructions* (Ziem / Lasch 2013; Goldberg 2006, 1995). In both these approaches, the dual role of a language user as the recipient as well as the source of linguistic conventions and change holds a central role. More recently, concepts such as *constructionalisation* (Traugott 2015; Traugott / Trousdale 2013; Hilpert 2008) or *cognitive sociolinguistics* (Hollman 2013; Grondelaers / Speelman / Geeraerts 2007) then fruitfully applied the method and concepts of constructionist approaches onto the discipline of diachronic language research, as well as sociolinguistic studies. At the same time, these advances show that, while rather comprehensive at every level of a linguistic system, construction grammar has a tendency to focus on more or less isolated constructions. Thus, it needs to remind itself that contextual factors like time in general, as well as a speaker's age, education or social class might influence the outcome of a study.

Still, as has been mentioned previously, creative alternations of collocations tend to be the exception rather than the norm, hence, not every speaker might use or even tolerate the same level of creative language. For the comprehension of complex syntactic constructions, for example, Dąbrowska[15] (1997) was able to show that native speakers' acceptance diverged drastically, based on the educational background of the participants. In Chipere (2003) as well as Dąbrowska and Street (2006), non-native speakers even outperformed their native speaker

14 Based on his conception of an ideal speaker-hearer Chomsky (1965: 4) argues that language falls into a normative – as Chomsky would argue – partly inborn system of rules, a speaker's *competence*, and his / her actual *performance*. This is not to be confused with Coseriu's (1973) dichotomy of *system* and *norm*, which distinguishes between potentially possible and actually used combinations of linguistic elements. Thus, *system* and *norm* could also be applied in a more usage-based and / or constructionist context, while *competence* and *performance* would contradict a view of language attainment where a speaker's system develops out of continuous language use.

15 In a later study, Dąbrowska and Street (2010) were able to obtain similar results in a comparison of participants with "LowAcademicAchievement (LAA)" and "HighAcademicAchievment (HAA)". In addition, they also showed that training resulted in a better performance for LAA test takers. Furthermore, in Dąbrowska (2010), participants' evaluations did not only vary depending on a participant's linguistic expertise, but also according to the degree of syntactic complexity.

counterparts in a task on the comprehension and recall of complex sentences and acceptability of plausible and implausible sentences. Thus, the connection between a collocation and its more creative alternations might be able to tell us something about the nature of collocations as such, and also serve as an indicator for potential stages of mental processing in language attainment. As they are generally constructed according to common syntactical patterns (like [Adj+N] or [VP+NP]) but apparently restricted combinatorially, collocations can neither be seen as a purely syntactical, nor as a clearly lexical phenomenon. They operate on an in-between level, which makes them a challenging subject for any comprehensive model of language.

However, since to date most studies have focused either on highly frequent combinations or collocations which were approved by academically trained evaluators[16], the status of more creative alternations ranges from deviant exceptions to creative instances of language use. Only rarely has creative language use been seen as a potential next step in an ongoing language development process. But in recent years, advances in usage-based theories and construction grammar have resulted in a new perspective on language acquisition and development. Seeing language as a continuum of ongoing change, several linguists began to approach language as a dynamic, complex adaptive system (Ellis / Larsen-Freeman 2009; Larsen-Freeman / Cameron 2008), which forms structures and abstractions from the input of its environment but also contributes to a steady change by feeding new and sometimes novel or creative utterances into the system. Thus, these approaches share the assumption that the human brain shapes structures and potentially, even rules, based on the input it receives. These structures are then combined to form new utterances. Here, studies in morphology (Bybee 1995) and syntax (Ambridge / Goldberg 2008; Tomasello 2005) have already been able to show that analogy plays a crucial role when it comes to the (re)combination of established items and structures, which might also be responsible for the development of novel and creative combinations. Yet, while morphology and syntax are traditionally regarded as a rule-based system with only a few idiosyncrasies, collocations are often defined by their idiomatic and unpredictable nature. As mentioned before, they could, therefore, be seen as in-between phenomena, which would make them interesting structures to fathom the interaction and effects of linguistic innovation and convention.

Therefore, this study seeks to investigate the constructional potential of collocations and also whether creative alternation within a collocational combination could lead to its manifestation as a construction. While context- and

16 Compare for example Howarth (1996), Nesselhauf (2004), Bartsch (2004) or Gyllstad (2007).

significance-oriented approaches alike have treated collocations as descriptively interesting exceptions, this text approaches the phraseological phenomenon from a (language) attainment point of view. Partly inspired by previous studies on *semantic prosody* (Stewart 2010; Sinclair 2004, 1991; Partington 2004; 1998; Bublitz 1995; Louw 1993), it argues that, to a certain extent, *collocations, creativity*, and *constructions* could also be interpreted as different stages of linguistic development. The following pages are based on the idea that, not unlike coins, which, in fact, are three-dimensional objects, collocations have often been approached from two angles: as a partly opaque, phraseological phenomenon, or as a rather interesting frequency effect (Granger / Paquot 2008; Bartsch 2004; Herbst 1996). A third dimension, consisting of temporal and social context, which, like the rim of a coin, might link both sides, has thus far gained little attention. This missing link could be found by looking at the genesis of individual collocational phenomena. But a closer examination of collocational creativity against a usage-based, constructional background is not just interesting from a purely theoretical point of view. Since English today is widely regarded as a key competence, not only in terms of vocational expertise but also as an essential skill in an increasingly private sphere which is becoming ever more global and international, the need to use and understand English correctly is still growing. In the last decades, a plethora of studies has already shown that, because of their seemingly unpredictable character, collocations are one of the most challenging phenomena for non-native speakers of English (amongst others: Howarth 1996; Granger 1998; Nesselhauff 2004; de Cock 2004). With seemingly arbitrary syntagmatic as well as paradigmatic restrictions, they operate outside the traditional "slot-filler model" (Sinclair 1991: 109). Against this background, items such as collocations, which do not allow for any random lexical filling within a syntactical structure, have to be memorized. Hence, even more advanced learners struggle with native-like English phrasing and might at times despair, since they might have been marked down for using combinations such as *pretty man* or *commit a mistake*, while native speakers seem to be allowed to use them, although admittedly only under certain circumstances. Thus, the three focal research questions (RQ) of this study are:

RQ1: Are collocations a cognitively stored entity, and if so, how can this perspective be adequately described in a comprehensive model of collocational combinations?

RQ2a: Is there a unified process underlying the attainment of collocational proficiency?

RQ2b: Does the collocational proficiency of native (L1) and non-native (L2) speakers of English develop in the same way?

RQ3: What role do the factors 'creativity' and 'context' play for the acceptability and analysis of collocational phenomena?

To approach these questions, the following pages deal with a more detailed discussion of the subject matter, collocations (> 2), as well as creativity and contextual change (> 3). Here the emphasis lies on the concepts themselves, and their theoretical background, as well as their position in modern (cognitive) language attainment research, to extrapolate which factors might contribute to the acceptance of creativity and whether these could be used to explain a collocation's creative alternation. Together with chapter 4, they contribute to RQ 1. Introducing a first cognitive conception of collocation (> 2) and discussing how and why change and creativity can be seen as a cognitive faculty (> 3), they lay the foundations for a more comprehensive model of collocation as a cognitive phenomenon within the process of language attainment (> 4). Thus, chapter 4 zooms in on the question of how approaches to language acquisition can contribute to RQ 1, and how these support a more comprehensive picture of the potential connections between collocations, creativity, and constructions. Combining different usage-based approaches towards phraseological phenomena (> 4.3), this *Dynamic Model for the Cognitive Development of Collocations* (DMCDC-model; > 4.4) is then put to a initial test in the subsequent chapters. In order to lay the methodological groundwork for RQs 2 and 3, chapter 5 is concerned with ways to operationalise the theoretical considerations from chapters 2 to 4 and discusses which measures need to be taken in order to be able to investigate differences between first and second language speakers as well as the role contextual factors might play. This methodology is then applied and discussed in chapters 6 and 7. Finally, chapter 8 lays out this study's major findings and implications. Figure 1.1 illustrates the outline of the present study.

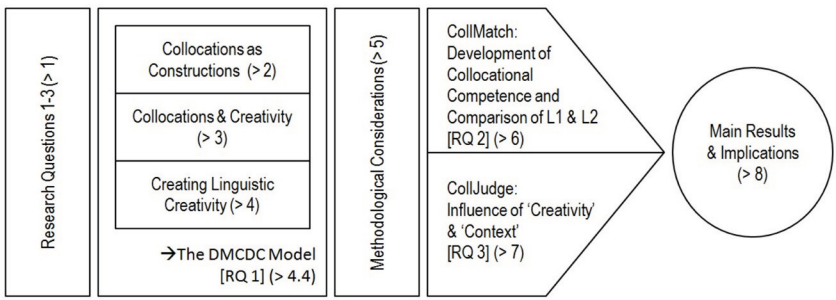

Figure 1.1: Scope of this Study

2 Collocations as Constructions

> When I use a word, 'Humpty Dumpty said in rather a scornful tone, 'it means just what I choose it to mean – neither more nor less.' 'The question is,' said Alice, 'whether you can make words mean so many different things.' […] 'When I make a word do a lot of work like that,' said Humpty Dumpty, 'I always pay it extra.' 'Oh!' said Alice. She was too much puzzled to make any other remark.
>
> (Carroll 1871 / 2001: 224–225)[1]

The debate about the defining properties of collocations seems to be as old as research within the field of lexical co-occurrences itself. Partly because studies on the character and value of collocations have been conducted for various reasons and purposes; the areas of research range from a very applied EFL context (de Cock 1999; Bahns 1997; Howarth 1996; Cowie / Howarth 1995; Hausmann 1984) to lexicography (Cowie 2012; Mel'cuk 1998; Benson / Benson / Ilson 1997; Bahns 1996; Hausmann 1985) and more theoretical, general linguistic description (Coseriu 1967; Firth 1951 / 1964, 1957 / 1968). However, a duality in conception might also have developed from the fact that interest in these special cases of lexical co-occurrence arose roughly at the same time within different schools of linguistics. British Contextualism, with its most prominent representative J. R. Firth at its centre, is frequently quoted as the cradle of the modern concept of collocation (Barnbrook / Mason / Krishnamurthy 2013: 36; Bartsch 2004: 30; Lehr 1996: 7), but also within lexicography authors like Palmer (1933), Hausmann (1984) or more recently Siepmann (2005) concerned themselves with lexical co-occurrences. Their focus still tends to lie more on the properties of collocations than on their contribution to the human linguistic system. Hence

1 In this quote from *Through the Looking-Glass*, first published in 1871, Carroll uses the sharp tongue of the egg-shaped creature Humpty Dumpty to remark on the polysemous character of words. Later, in the article "The stage and the spirit of reverence", the author picks up on the sociolinguistic dimension of this thought: "[…] no word has a meaning inseparably attached to it; a word means what the speaker intends by it, and what the hearer understands by it, and that is all … This thought may serve to lessen the horror of some of the language used by the lower classes, which it is a comfort to remember, is often a mere collection of unmeaning sounds, so far as speaker and hearer are concerned." (Carroll 1871 / 2001: 224). It is interesting to note that this quote not only emphasises the crucial role of textual as well as individual context, but also that utterances, which Carroll might have considered part of the "horror of some of the language used by the lower classes", today might be well established and accepted.

today, the field of collocational research seems to be split into contextual-oriented approaches and significance-oriented approaches[2] (Granger / Paquot 2008; Siepmann 2005; Herbst 1996). Based on John Sinclair's research (Sinclair 1991, 1966; Sinclair / Jones / Daley 1970 / 2005) on lexical frequency, contextual-oriented approaches nowadays mostly come in the shape of corpus-based, frequency-oriented research, while representatives of the significance-oriented approach focus on typological aspects relevant for the non-native language learner, such as (non-)compositionality and variability. Today, however, both sides seem to have reached a point where they realise that they have more in common than they disagree on.

Constructions, on the other hand, are in their broadest sense defined as "form and meaning pairings" (Goldberg 2006: 3). So, with the literal translation of collocation as a certain kind of "placing together" (Palmer 1933: 7), the term as such might suggest that the phenomenon of collocation is predominantly regarded as a formal or structural one, lacking an overall meaning dimension, which would be crucial for any kind of construction (> 4). But, as the following chapters will show, even very early accounts of collocation consider not only syntagmatic relations for the constituents of a collocation, but also discuss implications for meaning which stem from a contextual or paradigmatic level within an analysis. This suggests that collocations might be more than just formal, item-specific restrictions on word co-occurrences and that there is an inherent meaning dimension, which makes it possible to regard collocations as a form of construction in a construction grammar sense.

Following this idea, this chapter will use a selection of prominent approaches towards collocations from both camps and investigate their understanding of lexical co-occurrences to shed some light on their potential constructional character. Chapters 2.1 and 2.2 will, therefore, outline the two basic views on collocation and highlight potential connections to a modern construction grammar approach. Chapter 2.3 then concludes this section, suggesting a working definition and addressing some critical issues which need to be dealt with once a study assumes cognitive features to be part of collocational phenomena.

2 The following chapters will refer to these views as "context-oriented approaches" and "significance-oriented approaches". Often, "significance-oriented" is associated with statistical tests for significance, which would make "significance-oriented" a term to be attributed to corpus-based studies of collocations. The taxonomy of this study, however, is based on the respective focus of the different approaches. Thus, researchers who consider the context as the source of collocational combinations are grouped under the headline of "context-oriented", while definitions which focus on the prominence of an individual collocational combination are discussed under the label of "significance-oriented".