Collocational processing in typologically different languages, English and Turkish: Evidence from corpora and psycholinguistic experimentation

Doğuş Can Öksüz

This thesis is submitted in partial fulfilment of the requirements for the degree
of
Doctor of Philosophy in Linguistics
Department of Linguistics and English Language
Lancaster University

November 2019

Acknowledgment

I would like first to extend special thanks to my supervisors, Dr Patrick Rebuschat and Dr Vaclav Brezina who have been enormously supportive through the PhD process. Their academic knowledge and insight have been invaluable, and their assistance in identifying and taking advantage of opportunities has been second to none. I continue to be enormously grateful to have had the good fortune to work with them both.

I am also appreciative of Dr. Marije Michel, for sharing her thoughts early in the PhD process, and for her supervision of my work. I also would like to express my gratitude to Prof Morten Christiansen for his feedback, and Dr Aina Casaponsa for helping me with data analysis.

Many people in Turkey went out of their way to help me with data collection. There are too many to name, but special mention must go to my Uncle, Ruhi Çanak, for going far beyond the call of duty in wrangling volunteers to take part. I am indebted to him for his good-humour and his sanity. I also would like to thank Prof Belma Haznedar, Prof Yasemin Bayyurt, and Dr. Leyla marti for their enormous help with data collection at Bogazici University. I also would like to thank to the friends Merve Erdinç and Özlem Pervan and my cousin Ugur Kiliç who have been very supportive with data collocation.

To the friends who have remained very supportive and encouraging, I remain thankful and grateful for Saya Sato, and Emma McClaughlin.

I am deeply indebted to my partner Miranda Barty-Taylor (MBT) not only for proofreading and editing my thesis, but also being there for me whenever I needed her support.

Finally, but by no means least, thanks go to my parents, Dilek and Hüseyin Öksüz for taking a leap of faith in supporting me during this PhD. Without them I would not have had this chance. I dedicate this thesis to them. Sizlerin destegi olmadan bunu yapamazdim.

Abstract

Unlike the traditional words-and-rules approach to language processing (Pinker, 1999), usage-based models of language have emphasised the role of multi-word sequences (Christiansen & Chater, 2016b; Ellis, 2002). Various psycholinguistic experiments have demonstrated that multi-word sequences (MWS) are processed quantitatively faster than novel phrases by both L1 and L2 speakers (e.g. Arnon & Snider, 2010; Wolter & Yamashita, 2018). Collocations, a specific type of MWS, hold a prominent position in psycholinguistics, corpus linguistics and language pedagogy research. (Gablasova, Brezina, McEnery, 2017a). In this dissertation, I explored the processing of adjective-noun collocations in Turkish and English by L1 speakers of these languages through a corpus-based study and psycholinguistic experiments. Turkish is an agglutinating language with a rich morphology, it is therefore valid to ask if agglutinating structure of Turkish affects collocational processing in L1 Turkish and whether the same factors affect the processing of collocations in English and Turkish. In addition, this study looked at L1 and L2 processing of collocations in English.

This thesis firstly has investigated the frequency counts and associations statistics of English and Turkish adjective-noun collocations through a corpus-based analysis of general reference corpora of English and Turkish. The corpus study showed that unlemmatised collocations, which does not take into account the inflected forms of the collocations, have similar mean frequency and association counts in the both languages. This suggests that the base forms – uninflected forms of the collocations in English and Turkish do not appear to have notably different frequency and association counts from each other. To test the effect of agglutinating structure of Turkish on the collocability of adjectives and nouns, the lemmatised forms of the collocations in the both languages were examined. In other words, collocations in the two

languages were lemmatised. The lemmatisation brings the benefit of including the frequency counts of both the base and inflected forms of the collocations. The findings indicated that the vast majority (%75) of the lemmatised Turkish adjective-noun combinations occur at a higher-frequency than their English equivalents. In addition, agglutinating structure of Turkish appears to increase adjective-noun collocations' association scores in the both frequency bands since the vast majority of Turkish collocations reach higher scores of collocational strengths than their unlemmatised forms.

After the corpus study, I designed psycholinguistic experiments to explore the sensitivity of speakers of these languages to the frequency of adjectives, nouns and whole collocations in acceptability judgment tasks in English and Turkish. Mixed-effects regression modelling revealed that collocations which have similar collocational frequency and association scores are processed at comparable speeds in English and Turkish by L1 speakers of these languages. That is to say, both Turkish and English speakers are sensitive to the collocation frequency counts. This finding is in line with many previous empirical studies that language users process MWS quantitively faster than control phrases (e.g. Arnon & Snider, 2010; McDonald & Shillcock, 2003; Vilkaite, 2016). However, lemmatised collocation frequency counts affected the processing of Turkish and English collocations differently, and Turkish speakers appeared to attend to word-level frequency counts of collocations to a lesser extent than English speakers. These findings suggest that different mechanisms underlie L1 processing of English and Turkish collocations. The present study also looked at the sensitivity of L1 and L2 advanced speakers to the frequency of adjectives, nouns and whole collocations in English. Mixed-effects regression modelling revealed that L2 advanced speakers are sensitive to the collocation frequency counts like L1 English speakers because as the collocation frequency counts increased, L1 Turkish-English L2 speakers responded to the collocations in English more quickly, as L1 English speakers did. The results indicated that both groups showed sensitivity to noun frequency counts, and L2 English advanced speakers did not appear to rely on the noun frequency scores more heavily than the L1 English group while processing adjective-noun collocations. These findings are in conflict with the claims that L2 speakers process MWS differently than L1 speakers (Wray, 2002).

Contents

List of Appendices

List of Tables

List of Figures

Introduction

The ability to express our thoughts using an infinite number of linguistic means is one of the most well-known features of human language. From this perspective, words and rules have previously been viewed as finite means and combining them enables us to reach an infinite linguistic productivity. In other words, words and rules were viewed as the fundamental building blocks of language (Pinker, 1999), and multi-word sequences (MWS) were sidelined as rare expressions. This words-and-rules approach, also known as generative grammar, adopts an atomistic view of language and suggests that knowledge of language is based on a system of rules used to generate a large number of structures (Chomsky, 1965, pp 15-16). This view of language relies on the assumption that language users employ the creative syntactic rules to a full extent. However, an increasingly widespread assumption is that MWS, rather than the individual words are the building blocks of both written and spoken discourses (Altenberg, 1998; Christiansen & Chater, 2016a; Goldberg, 2006; Granger & Meunier, 2008; Pawley & Syder, 1983; Sinclair, 1991; Wray, 2002). MWS can be loosely defined as fixed or semi-fixed recurrent phrases (Siyanova-Chanturia & Martinez, 2014). They include collocations, *strong tea*; binomials, *black and white*; discourse markers *by the way;* idioms, *spill the beans*; proverbs, *better late than never*; speech formulae, *what's up*; lexical bundles, *in the middle of*; and so on. These sub-types of MWS reflect varying degrees of semantic non-compositionality, syntactic fixedness, lexical restrictions, and institutionalization (Evert, 2005; Granger & Meunier, 2008; Wray, 2002).

Evidence from various corpus analyses suggests that language appears to be more repetitive than it was predicted to be by the words-and-rules approach (e.g. Pinker, 1999), and recurring MWS are far from being rarely used expressions. For example, Jackendoff (1997) estimates that MWS appear as many as single words in everyday language use and concludes that “they are hardly a marginal part of our use of language” (p. 156). In line with this finding, DeCock, Granger, Leech, and McEnery (1998) estimate that MWS constitute a large proportion (up to 50%) of the language produced in both written and spoken discourses by native-speakers. Furthermore, Erman and Warren (2000) found that 52.3% of the written discourse they looked at consisted of MWS. In other words, language users seem to know and use many recurring MWS. The study of MWS, also known as formulaic language, is based on the insight that some sequences of words which could potentially be analysed into smaller units are considered to be fixed sequences and tend to be treated as wholes (Durrant, 2013). In some cases, sequences are treated as wholes because their meaning and syntactic behaviour is not predictable from a more general knowledge of language such as idioms (e.g. *kick the bucket*). In other cases, sequences are treted as wholes because although they are semantically and syntactically regular, they have been accepted by the speech community as the conventional way of expressing a particular message (e.g. *long live the king*). In the case of collocations, a specific form rather than a meaning-equivalent one is adopted by the speech community, for example, the adjective-noun collocation *fatal mistake* is considerably more frequently used than the meaning equivalent adjective-noun pair *extreme mistake*. The choice of one form over another is largely arbitrary (Durrant, 2013), and nativelike collocational sensitivity requires specific knowledge of such forms (Pawley & Syder, 1983). The frequent use of some forms rather than their meaning-equivalent ones is also associated with some form of processing efficiency (Goldberg, 2006).

The study presented in this thesis is mainly concerned with the mental processing of MWS in typologically different languages. More specifically, it attempts to investigate whether collocations in typologically different languages are processed similarly or whether different factors affect their processing in different languages. This study also explores the mechanisms affecting the processing of MWS in speakers' first (L1) and second languages (L2). As Biber (2009) and Durrant (2013) also have recently noted, agglutinating languages such as Turkish and Finnish provide particularly interesting area for exploration of MWS. These languages employ extensive system of suffixes to build up complex word forms. Therefore, the study of MWS in agglutinating languages is interestingly different fron non-agglutinating languages such as English and Dutch. Their rich morphology increases the possibility that complex types of formulaicity could be found within as well as between words. It should also be noted that definition of formulaicity have recognised that formulaicity can be found in linguistic units at all levels (Wray, 2008). There has been only little research focusing on formulaicity in agglutinating languages. Some psycholinguistic studies investigated the roles of rules and memory in the processing of morphologically complex words in Finnish (Lehtonen & Laine, 2003; Soveri, Lehtonen, Laine, 2007). Furthermore, Durrant (2013) carried out a corpus-based investigation to provide a detailed description of formulaic patterns within Turkish words. He found that many prototypically formulaic patterns take place at the morphological level in Turkish. Building on this finding, this study firstly explores whether agglutinating structure of Turkish affect the formulaic patterns between words. More specifically, this study looks at collocations occur at similar frequency levels in the two languages. Secondly, this study explores whether the same or different factors play role in the processing of MWS for L1 speakers of these languages.

This introductory part provides the background of the study. It starts with a section that presents my motivation behind the study (0.1). Section 0.2 briefly explains the main aims and the research design. Section 0.3 presents the layout of the thesis.

Motivation

The motivation behind this study comes from my desire to broaden the scope of the study of MWS by incorporating typologically different languages. As has been acknowledged by Biber (2009), and Durrant (2013), the research on MWS is limited to a few European languages, especially in English. Therefore, broadening the range of languages potentially contributes to the understanding of formulaicity, particularly with regard to how language typology affects the mental processing of MWS. I also used this study as an opportunity to combine corpus and experimental methods to carry out a contrastive study on processing of MWS in English and Turkish. Since the methodologies of corpus and psycholinguistic experiments have complementary strengths and they can provide insights about different aspects of language use and processing (see section 2.1 for integrating corpus and experimental methods), this combination should be fruitful for exploring the processing of collocations in typologically different languages. Another motivation for me to conduct this study was based on a few observations as a language learner and teacher. I had had the chance to observe that second language learners especially adult learners experience difficulty using collocations accurately and fluently. In this study I wanted to contribute to the existing body of knowledge with respect to whether L1 and L2 speakers process the collocations in fundamentally different ways.

Aims and the design of the thesis

Biber (2009) suggests that future research on MWS should involve languages other than English, especially languages with different typological profiles. As Durrant (2013) also notes, an important weakness in this field of study is that the previous studies predominantly focussed on a narrow range of languages, and particularly on English. As a result, the status of MWS as a general feature of language is not yet well established yet. Therefore, the present study aims to address this gap in two ways. This thesis firstly presents findings of a corpus study that comparatively investigates English and Turkish collocations' frequency of occurrence, and collocational strength. Turkish is an agglutinating language with a rich-morphology. Thus, it is interesting to observe if the agglutinating structure of Turkish affects the collocability of the words. The goal of this corpus study is to observe the extent to which Turkish and English collocations are comparable in terms of frequency and collocational strength; important features of collocations (see Brezina, Wattam, & McEnery, 2015 for a review on collocational properties). The present study secondly examines the mental processing of collocations by L1 speakers of English and Turkish. More precisely, the findings of the experiments will provide evidence regarding the extent to which L1 of English and Turkish rely on the same mechanisms (such as single word and collocation frequency counts) for the processing of high- and low-frequency collocations. The current study also aims to observe the extent to which the findings of the experimental work align with the results of the corpus study.

The current study also examined whether L1 and advanced level L2 speakers of English process high- and low-frequency collocations in qualitatively different ways. It has been argued by some researchers, most notably by Wray (2002, 2008) that L1 and L2 speakers process MWS in qualitatively different ways. That is to say, L1 speakers rely on their knowledge of meaning assigned to MWS whereas L2 speakers decompose MWS into their individual components and rely heavily on words and rules to process them. However, some key studies on collocations did not provide empirical support for this position. For example, Durrant and Schmitt (2010) found a learning effect for collocations presented in a training session to adult L2 learners and suggested that it is the lack of exposure to L2 input that explains the differences between L1 and L2 collocational processing. In line with this finding, Wolter and Gyllstad

(2013) investigated advanced L2 speakers' processing of adjective-noun collocations and found that advanced L2 speakers of English showed sensitivity to frequency effects for collocations. More recently, Wolter and Yamashita (2018) also found that both NSs and NNSs showed sensitivity to both collocational and single word-level frequency counts. The present study partially replicates these experimental works (e.g. Wolter & Yamshita, 2018) by examining the processing of high and low-frequency collocations by L1 and advanced level L2 speakers of English. The present study aims to find out the extent to which single word and collocation frequency scores affect the processing of high and low-frequency collocations by L1 and L2 speakers of English.

Structure of the thesis

This PhD thesis consists of four main parts. In Part 1, chapter 1 discusses the theoretical positions which attempt to explain the psycholinguistic reality of MWS (e.g. lexical priming by Hoey, 2005), Chapter 2 firstly discusses the ways of successfully combining corpus and experimental approaches. It secondly introduces the current study including the research questions. In part 2, chapter 3 reviews the previous literature in corpus studies on MWS. It firstly discusses the collocational properties and corpus-based collocational association measures. It secondly reviews the contrastive corpus studies on MWSs in different languages. Finally, it reviews the corpus studies on MWS in agglutinating languages such as Turkish. Chapter 4 reports on a comparative corpus study of collocations in English and Turkish. The corpus study reported in this thesis aims to compare collocations' frequency and collocational strength in English and Turkish. In part 3, chapter 5 reviews the previous literature of experimental work on the processing of MWS by L1 and L2 speakers. Chapter 6 reports on psycholinguistic experiments examining the processing of collocations in English and Turkish by L1 speakers of these languages. In addition, it investigates the processing of collocations by L1 and advanced L2 speakers of English. In part 4, chapter 7 provides a general discussion on the findings of corpus and experimental works. It specifically focusses on the processing of collocations in typologically different languages English and Turkish, and the processing of collocations by L1 and L2 speakers. Furthermore, it discusses the benefits and challenges of combining corpus and experimental methods.

Part 1: Theoretical and Methodological Frameworks of the Study

This part discusses the theoretical positions about mental processing and representation of MWS (see Chapter 1), and it discusses the ways of combining psycholinguistic and corpus methods; the methodological approach of the current study (see Chapter 2). Chapter 1 firstly provides definitions for major types of MWS (e.g. collocations, idioms). It secondly summarises theoretical approaches foucsing on how MWS are processed. Some of these theoretical positions adopt a dualistic-paradigm in which phrases are either stored holistically or processed morpheme-by-morpheme (see sections 1.3, 1.4, 1.5). A common feature of these approaches is that they propose a sharp distinction between the MWS, which are stored holistically in the mental lexicon, and the analytical system that carries out grammatical processing of the single words (see e.g. Sinclair, 1991; Wray, 2002). It secondly presents more recent theoretical approaches regarding the processing of MWS such as Hoey's (2005) lexical priming and the usage-based approaches to language acquisition (see 1.4 and 1.5). The usage-based approaches (see section 1.7) to language acquisition rejects a sharp distinction between holistic and analytical systems, and they perceive an important connection between the mental processing and representation of linguistic items and speakers' lifetime experience with the language (Christiansen & Chater, 2016; Kemmer & Barlow, 2000; Ellis, 2002; Tomasello, 2003). In this section (1.7), a special attention has also been paid to "the implicit statistical learning" (e.g. Saffran, Aslin & Newport, 1996) and the "now-or-never bottleneck" frameworks (Christiansen & Chater, 2016b), which provide plausible and testable arguments for why MWS as well as the single words are the building blocks of both written and spoken discourses (see sections 1.7.1 and 1.7.2). Chapter 2 discusses the different ways of combining corpus and experimental methods to study the cognitive processing and use of MWS (see section 2.1). It then provides some terminological clarifications about MWS, since various terms have been used to refer to the similar constructs in a potentially confusing way. Finally, it presents the current study reported in this thesis (see section 2.3) including the research questions that are investigated in the following parts (see parts 2 and 3).

Chapter 1: Reviewing Theoretical Approaches to the Mental Processing and Representations of MWS

Recently MWS have received considerable attention in psycholinguistics, applied linguistics and corpus linguistics. The reason is that they are both very frequent and appear to create a processing advantage for language users (Siyanova-Chanturia & Martinez, 2015; Carrol & Conklin, 2019). MWS vary in terms of structure, completeness, length, and transparency of meaning, however empirical studies that looked at the processing of various types of MWS have found relatively consistent results (Siyanova-Chanturia, 2015; Siyanova-Chanturia & Martinez, 2015). That is to sayl, a vast body of psycholinguistics studies found that frequent MWS are processed faster than the control phrases (e.g. Durrant & Doherty, 2010; Siyanova-Chanturia, Conklin, & Van Heuven, 2011; Tremblay, Derwing, Libben & Westbury, 2011; Vilkaite, 2016; Yi, 2018, see chapter 5 for a detailed review of these studies). On this basis some researchers have even claimed that MWS are stored as wholes - and accessed as wholes when needed - in the mental lexicon (e.g. Wray, 2002; 2008). Therefore, they were processed faster than the control phrases. On the other hand, other researchers have argued that the MWS are processed faster than the novel phrases because language speakers are sensitive to the phrasal frequency of the MWS and this is in line with the usage-based models of language acquisition (Barlow & Kemmer, 2000; Ellis, 2002; Tomasello, 2003). This chapter therefore, firstly focuses on the main characteristics of MWS and then it reviews the theoretical approaches which attempt to explain why MWS are processed faster than the novel phrases and why they are a pervasive feature of language.

1.1 The definition and characteristics of MWS

Words tend to co-occur in specific linguistic configurations, which are known as MWS (Carrol & Conklin, 2019; Siyanova-Chanturia, Conklin, Caffara, Kaan, van Heuven, 2017). The study of MWS has played a prominent role in the study of language learning and processing for several decades (Wray, 2013). Published reseach on MWS has cut across the fields of psycholinguistics and corpus linguistics and language teaching (Gablasova et al. 2017a) and identified a broad range of constructions (e.g. collocations, lexical bundles, idioms, binomials). They fulffill a number of communicative functions (Wray, 2002, 2008), and knowledge of each type of such constructions play prominent roles in how we use language. For example, in English we unconsciously ask for *salt and pepper*, but not *pepper and salt* and use the adjective *strong* but not powerful to describe coffee. Two features are common to all types of constructions that are MWS. The first is that they are recurrent. That is to say, they occur more frequently than comparable novel phrases in a natural language. However, what counts as the threshold for frequency is an open question and different theoretical positions have different approaches to frequency of MWS compared to novel phrases. The second feature is that they are processed faster than novel control phrases (Carrol & Conklin, 2019). In other words, MWS have frequency effects at multiword level that is in line with the usage-based based approaches to language acquisition (see section 1.6 for a detailed discussion on usage-based approaches). Beyond this very broad designation, MWS vary along a number of important linguistics features (Titone, Columbus, Mercier, Libben, 2015), such as their degree of fixedness/ conventionalisation, their semantic unity, their degree of compositionality and the function they perform (see also Buerki, 2016 for an overview). At this point, it is useful to look at different types of commonly studied MWS – idioms, binomials, lexical bundles and collocations – that all must be considered as MWS from the point of view of frequently occurring word sequences.

Idioms are among the most studied of all types of MWS and have been described as prototypically formulaic (Siyanova-Chanturia & Martinez, 2015). This is probably because they carry many features of MWS such as high-frequency, fixedness, compositionality, and semantic unity (Titone et al. 2015). They are semantically opaque, self-contained figurative phrases such as *kick the bucket.* In addition to being a clear example of MWS, idioms are considered as examples of figurative language, which may play a role in how they are processed. A robust finding in the literature is that idioms are generally recognised more quickly than the matched novel phrases. For example, Swinney and Cutler (1979), using a phrasal decision task, found that idioms (e.g. *break the ice*) are recognised as meaningful phrases faster than the control phrases (e.g. *break the cup*). Recently, eye-tracking experiments also provided empirical support to the processing advantage of idioms over control phrases (Carrol & Conklin, 2017), regardless of whether they are used in figurative or literal contexts. The literature suggests that the processing advantage of the idioms stem from the fact that they are highly familiar (Schweigert & Moates, 1988, and predictable (Libben & Titone, 2008).

Binomials, one of the most commonly studied MWS, are defined as combinations of *x-and-y* (e.g. *fork and knife*) where a reversal of the order is entirely possible, but where on word order is conventionalsed (Carrol & Conklin, 2019). The most common examples are noun-and-noun (e.g. *salt and pepper*). Many factors were found to affect the word order of the binomials such as conceptual factors (general before specific), cultural restrictions (power relations), and phonological variables (e.g. Benor & Levy, 2006; Lohmann, 2012; Mollin, 2012; Morgan & Levy, 2016). Frequency of binomials has found to affect their processing (Conklin & Carrol, 2016; Siyanova-Chanturia, Conklin, van Heuven, 2011). Siyanova-Chanturia et al. (2011) argued that phrasal frequency only partially affects their processing because binomials (e.g. *bride and groom*) are processed faster than their reversed forms (e.g. *groom and bride*). They proposed that the word order itself played an important role in their processing. Since their preferred word-order enjoyed processing advantage over reversed forms, even when the phrasal frequency was controlled (see section 5.1 for a more detailed review of processing binomials).

Lexical bundles are also relatively common type of MWS. They may span phrasal boundaries. Biber and Conrad (1999) defined lexical bundles as sequences of two, three or fours words that occur as wholes at least 10 times per million words. Some examples that fit this definition are that *I don't know whether*, *on the other hand*, and *in front of the.* Lexical bundles do not have idiomatic meaning and fulfil a wide range of discourse functions. According to Biber, Conrad and Cortes (2003) they can be used as referential units, text organisers, or interactional units. Lexical bundles are structurally complex, they are usually incomplete and not fixed. Common types of lexical bundles also include noun phrase and prepositional phrase fragments such as *one of the most*, *an increase in the* and *in the light of*. In line with other types of MWS, high-frequency lexical bundles are processed faster than control phrases (Bod, 2001; Jiang & Nekrasova, 2007; Tremblay, Derwing, Libben, Westbury, 2011).

Collocations are also common type of MWS. They can be broadly defined as frequently co-occuring word sequences (see section 2.2.1 for a specific definition of collocations). Typically, they may entail verb-noun (e.g. *play a role*) and adjective-noun pairs (e.g. *fatal mistake*). They can also be partially idiomatic (e.g. *draw a conclusion*), and various classifications from a phraseological point of view have been proposed (see section 2.2.1 for a more detailed discussion on the operationalisation of collocations). There is a clear evidence that frequently occurring collocations are processed faster than control phrases (Durrant & Doherty, 2010; Sonbul, 2015, Vilkaite, 2016, see sections 5.1 and 5.2 for a detailed review of literature for collocational processing). There is a robust evidence in the literature that frequency plays an important role in the processing of collocations, but other factors such as transitional probabilities as measured by association measures (e.g. Delta P, Mutual information) seem to affect their processing. The following sections (1.2, 1.3, 1.4, 1.5, 1.6 and 1.7) present the theoretical positions focussing on the on-line processing and representation of MWS.

1.2 The words-and-rules approach

For some researchers, the fact that particular word seqeuces are considerably more frequently used by native-speakers in preference to alternative, equally grammatical sequences should be explained by the realm of language use or in Chomky's (1965) terminology, linguistic performance. Researchers working in this theoretical framework views linguistic competence as knowledge of a generative grammar. In other words, the linguistic phenomenon of formulaicity does not influence the way in which language competence should be defined. As argued by Pinker (1999), using pre-fabricated chunks of language is a peripheral aspect that is not related to real language processing (p. 90). More recently, Pinker and Ullman (2002), and Ullman (2001) extended the words-and-rules approach to a neurocognitive substrate of lexicon and grammar. According to Declerative/Procedural hypothesis (Ullman, 2001), declerative memory, which involves lexical memory, stores facts, events and arbitrary relations. The declerative memory, rooted in temporal lobe structures, appears to be responsible for associative binding and underlies not only the learning and use of facts and events and but also of the sounds and meanings of words, that is also known as the mental lexicon (Ullman, 2001). Ullman (2001) further suggests that the mental lexicon may store some distinctive information that is smaller or larger than words such as bound morphemes (e.g. *ed* or *-ness*) and representation of complex linguistic features whose meanings cannot be transparently derived from their parts such as idiomatic expressions (e.g. *kick the bucket*). The procedural memory rooted in frontal/basal ganglia structures, may be specialised for sequences and underlies the implicit learning of motor and cognitive skills and grammatical rukes in both syntax and morophology. In addition, Ullman (2001) suggests that even if a mental rule exists for a given transformation, the linguistic items cqan be learned and subsequently computed in associative memory. In this case, a given linguistic item or form can be computed either in associative memory or by the rule system. The frequency of the items has an influence on the likelihood of storage and retrieval of them in the associative memory. As the frequency of the forms and items increase, they are more likely to be stored and retrieved from the associative memory. It should be noted the extended version of the words-and-rules appears to be more compatible with MWS and the effect of frequency than its earlier version. However, it does not spicifally explain how MWS are processed and represented. The next sections (1.2) and (1.3) summarise the theoretical positions which specifically focus on how MWS are processed and repressented.

1.3 The puzzles of nativelike selection and nativelike fluency

Pawley and Syder (1983) may have been one of the first sets of researchers who explored the processing and representation of MWS. In their seminal paper, they proposed a model of linguistic competence that puts emphasis on the psycholinguistic reality of MWS. The model consisted of two components; the puzzle of "nativelike selection" and the puzzle of "nativelike fluency. They defined "nativelike selection" as the ability to select natural and idiomatic phrases from a wide range of grammatically correct expressions and "nativelike fluency" as the ability to produce connected speech. The "nativelike selection" is particularly relevant to the use of MWS. They argued that the main weakness of "the words-and-rules approach" (Pinker, 1999) is to assume that native-speakers enjoy the creative potential of syntactic rules to their full extent. Furthermore, they suggested that the presence of a large proportion of MWS in language blurs the distinction between lexicon and grammar. According to Pawley and Syder's (1983) model, L2 learners should not only learn the words and rules of grammar, but also acquire a means for identifying which of the grammatical expressions sound natural, nativelike (Pawley & Syder, 1983, p. 183). Although, they did not appear to provide an operational definition of the notion of 'naturalness', their contribution was very important for acknowledging that language competence is not limited to mastering the rules of grammar, and grammatical competence should be viewed as only one of the components of native-like competence. The next section (see 1.4) summarises Sinclair's idiom principle approach which provides a corpus-based perspective to the processing and use of MWS.

1.4 The idiom principle

Large corpora of authentic language data provide a very suitable method for studying MWS (Bartsch & Evert, 2014). Sinclair is one of the pioneer researchers, who developed corpus-based methods for observing recurrent patterns of MWS across large text collections. Corpus analysis clearly suggest that a single word should not be viewed as a unit of meaning by default (Sinclair 1991; 2004; Stubbs, 2009). That is to say, meaning cannot necessarily be conveyed by a single word. A debate emerged out of this corpus evidence about the extent to which the meaning of the individual words differs when they are part of MWS - which is viewed as a unit of meaning. Previously, the assumption was that a word could inherently have one or more than one meanings. However, corpus-based phraseological analyses enabled the researchers to notice that the surrounding words or MWS in which words appear to outweigh the number of meaning a word possibly has because many meanings require more than one word for their normal realizations (Sinclair, 2004: 132). It has also been observed that the use of individual words contributes to the meaning in a way which is not related to their meanings in a dictionary

(Sinclair, 2008b: 408). For this reason, learning the literal meaning of a word without leaning its phraseology does not suffice for using this word accurately in a context (Stubbs, 2009). As a result, units of meaning should be much more extensive and varied than has been observed in a single word (Cheng, Greaves, Sinclair, & Warren, 2009). The idiom principle suggests that form and meaning cannot be easily distinguished because a grammatical meaning is generated by a choice and a meaning cannot exist without a choice (Sinclair, 2004, 2008).

Sinclair (2004) argued that the tradition of linguistic theory favoured the paradigmatic choice rather than the syntagmatic one. That is to say, the texts were considered as relatively independent choices of lexical items, and the patterning was seriously downplayed. Although, the individual lexical items were characterized as they have hundreds of potential meanings which were not likely to be available, because they are constrained by the other choices in the immediate context. The major reason for this biased view of language description was the unavailability of the corpus approaches and the technology associated with it, which has since enabled the researchers to statistically observe how different linguistic choices co-ordinate with each other, and share the realization of meanings across texts (Sinclair, 1991 p. 78, Stubbs, 2009 p. 121). The balanced view of syntagmatic and paradigmatic choices would provide a view of the units of meaning, which is characterized as the most invariable form. Vetchninikova (2014) suggest that conjunctions such as *but, and, however* form their own independent units of meaning, however, many frequent verbs such as *make,* or *take* can rarely form their own isolated meaning units without patterning with neighbouring words.

Sinclair (1991) argued that meanings are constructed out of two distinct processes, which are the open-choice and idiom principles. The open-choice principle, on the one hand, is based on the assumption that a large range of choices opens up to be filled by individual words on a quite random basis and grammaticalness is the only constraint. This is also defined as the "slot and filler" model; any word, which does not violate the rules of grammar, can occur in any slot. This principle operates on the paradigmatic choice. On the other hand, Sinclair (1987, 1991) suggested that most words used in daily language, do not have an independent meaning, but are the component of a rich repertoire of multi-word patterns or MWS. Therefore, a typical text is not produced simply out of open-choice principle. This argument is in line with a number of corpus analyses emphasizing that a certain proportion of typical texts consist of MWS (Altengberg, 1998; Erman & Warren, 2000; DeCock et al. 1998) The idiom principle lies at the heart of Sinclair's conceptualisation of how language users process the stretches of language. The definition of the idiom principle is as follows:

> The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments. (Sinclair, 1987: 320; 1991: 110)

Primarily, the idiom principle suggests that those co-selected words operate as a single word (Sinclair, 1991 p. 110), although they seem to be decomposable into smaller segments. In this regard, Sinclair (1991) claimed that the meaning of those co-selected words are holistically stored by language users and thus words cannot remain independent of the other words unless they are either very rare or specifically protected (Sinclair, 2004: 30). In line with this, Vetchinnikova (2014) pointed out that co-selection and delexicalisation are the components of one single process. What is co-selected is delexicalised and consequently delexicalisation leads to the meaning shift. Delexicalisation is when a word participartes in a unit of meaning its core meaning vanishes. In other words, the delexicalisation applies to the co-selected words only (Cheng et al., 2009; Vetchinnikova, 2014). Furthermore, Sinclair (2004) proposed the terms "terminological tendency" and "phraseological tendency". Terminological tendency, defined as the tendency for a word to have a fixed meaning in reference to the world, that highlights the open-choice principle. Phraseological tendency, defined as the tendency towards idiomaticity, (words' getting together and making meaning together through combination), highlights the idiom principle. Sinclair (2004) suggested although a preponderance of language use lies between these two tendencies, there are very few invariable, fixed phrases highlighting the terminological tendency in English, and this is an evidence for the key role of phraseology in the description of English. In line with Sinclair's perspective, Wray's (2002, 2008) also offers a dualistic model of processing for MWS.

1.5 Wray's dual system approach to MWS

In relation to MWS, Wray (2002) adopted the term "formulaic sequences", from various perspectives including processing in L1 and L2. Wray's (2002) model proposes a dual-system, consisting of holistic and analytic systems. Wray (2002) argued that the advantage of the analytic system is that it is flexible for comprehending and producing novel input. The advantage of the holistic system is that it reduces speakers' processing effort. It is simply more efficient and effective to retrieve a prefabricated string than to create a novel one. Therefore, communicative competence does not only involve knowing rules for the composition of sentences and using these rules to create expressions from scratch, but it also involves knowing an inventory of prefabricated expressions and a kit of rules to be able to make the necessary adjustments according to the contextual demands. In this regard, communicative competence also involves the adaptation of prefabricated patterns to the contextual demands and the rules are regulative and subservient rather than generative (p. 18). Therefore, Wray (2002) views MWS as having a privileged processing status and she seemed to interpret this as an evidence for holistic storage. Wray (2002) came up with the definition of the MWS below, formulaic sequences in her terms.

> (A formulaic sequence) is a sequence continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is stored, retrieved, wholly from memory at the time of use, rather than being subject to generation or analysis by the language grammar (Wray, 2002: 9).

As Wray (2002) acknowledges, this definition aims to be as inclusive as possible since it covers a wide range of linguistic units that can be considered formulaic (e.g. lexical bundles, collocations, idioms). In this definition she strongly argued that MWS are prefabricated because they are retrieved as a whole from the lexicon without any grammatical processing taking place. Furthermore, she seems to take a step further proposing that different sizes of MWS and internal structure can be stored in the lexicon as "morpheme equivalents" (Wray, 2002 p. 266). That is to say, instances of MWS are stored and retrieved whole from memory rather than being subject to analysis by the language grammar (Wray, 2002) that they behave akin to a single morpheme (Wray, 2008). This proposition was further reviewed in Wray (2008), where she further defined the concept of a "morpheme equivalent unit".

> Morpheme equivalent unit is a word or word string whether incomplete or including gaps for inserted variable items that is processed like a morpheme, that is, without recourse to any form-meaning matching of any sub parts it may have (Wray, 2008: 12).

The key tenets of the both of these definitions are that MWS are stored and retrieved whole from the memory rather than being subject to analysis by the language grammar (Wray, 2002), and that they behave similarly to a single morpheme (Wray, 2008). Put simply, when we encounter a highly frequent phrase such as *fast food* we do not necessarily activate or access the individual components of the phrase *fast* and *food* in our mental lexicon. In line with Wray's (2002, 2008) view, Sinclair (1991) also argued that the language users holistically process the semi-preconstructed phrases that constitute "single choices". Furthermore, Pawley and Syder (1983) claimed that speakers are able to retrieve the MWS as wholes or as automatic chains from their long-term memory. However, these researchers' conceptualisations of such holistic processing or retrieval of MWS is not entirely clear. Pawley and Syder's (1983) wording "retrieved as a whole" implies that the MWS, are stored and retrieved holistically. However, their expression "retrieved as automatic chains" implies that MWS simply facilitate processing without making the claim that they are processed holistically. Similarly, Sinclair's (1991) notion of "single choices" does not seem to be very clear. In his definition, the notion of "single choices" appears to semi-preconstructed phrases that can be decomposed into their constituent parts. This in turn raises questions about the degree of lexical fixedness of MWS, and the extent to which the language speakers handle the word insertions when they process MWS. Moreover, all these researchers argue that frequency plays a role in holistic storage; that is, MWS of sufficient frequency can attain independent representation as a way of making processing more efficient (Sinclair, 1991; Wray, 2002). In this frequency threshold approach, there is a distinction between MWS that are stored holistically, and the ones that are not. Nevertheless, they do not elaborate on any specific frequency thresholds at which we should expect holistic storage, or how we can calculate the frequency thresholds for different types of MWS.

Wray (2002, 2008) also argued that L1 and L2 speakers process MWS in fundamentally different ways. That is to say, L1 speakers rely on their knowledge of meaning assigned to MWS whereas L2 speakers decompose MWS into individual words and rely heavily on words and rules to process them. Wray (2002) suggests that L1 speakers would simply know that idiomatic way to talk about a big disaster is to call it a *major catastrophe* and L1 speakers would process this phrase as a single unit of meaning. In contrast, she suggests that L2 speakers would break this into its individual words and process them through the semantic value of the component words (p. 209). This view has not received much empirical support. Durrant and Schmitt (2010), for example, found a learning effect for collocations presented in a training session, and suggested that it is the lack of exposure to L2 input that explains for the differences between L1 and L2 processing of MWS. Along these lines, advanced L2 speakers were found to be sensitive to frequency effects for collocations in online tasks (Wolter & Gyllstad 2013, Wolter & Yamashita, 2018; Siyanova-Chanturia, Conklin, & Van Heuven, 2011 for a detailed review of these studies see chapter 5). Furthermore, these findings were also supported by a meta-analysis carried out by Durrant (2014) who found positive correlations between frequency counts of collocations and learner knowledge of collocations.

In the following sections (see sections 1.6, and 1.7), firstly Hoey's (2005) lexical priming approach has been presented (see 1.6). Secondly usage-based approaches to language acquisition are reviewed (see 1.7).

1.6 Lexical priming

Hoey's (2005) account of lexical priming is concerned with the questions of how naturalness is achieved, that is ability to select natural and conventional word sequences from a wide range of grammatically correct expressions and how an explanation of what is natural in a language influences what is possible in a language. Hoey (2005) defined collocations as a psychological association between the words, evidenced by their co-occurrence in corpora more frequent than random distribution can explain. For instance, co-occurrence of the words; *in* and *winter* is much more frequent than that of the words; *through* and *winter* (Hoey, 2005 p. 6), so that the preposition *in* is more likely to prime *winter* than the preposition *through*. In this regard, Hoey (2005) pointed out that every word is mentally primed for collocational use and thus language users' implicitly draw on the collocations to produce meaningful sentences. Furthermore, he outlined the notion of priming as language users' having access to a mental concordance, which is richly glossed for social, physical, discoursal, generic and interpersonal contexts, for every word they encounter. That is to say, every word is primed to occur within the company of particular other words, which are its collocates (Hoey, 2005 p. 13). Furthermore, Hoey (2005) provides evidence that priming is not only limited to lexical items. Words can also be primed to occur in certain grammatical categories and with certain semantic sets. For instance, the word pair *in winter* has a tendency to be reported predominantly in the present tense. In addition to the grammatical constraints, words can also prime certain semantic sets. For instance, the words *actor* and *actress* collocate with *director*, *best*, *film*, *singer* and *former*. However, the word *architect* collocates with a completely different semantic set; *designed*, *new* and *chief*. (Hoey, 2005). That is to say, different terms of occupations such as *architect* and *actor* are primed quite differently from each other. Nevertheless, both of these words share a large number of functional words as their collocates such as definite and indefinite articles, possessives or a post modifying *of*-phrase such as *the architect's brief, the skills of an architect*. In the next section (see 1.5), the usage-based approaches to language acquisition, the implicit statistical learning and the now-or-never bottleneck frameworks are presented.

1.7 Usage-based approaches to language acquisition

According to usage-based approaches to language acquisition, speakers' language systems are intimately shaped by their lifetime experiences of the language (Bybee, 1998; Ellis, 2002; Goldberg, 2006; Kemmer & Barlow, 2000; Tomasello, 2003) which hold that speakers learn constructions while they engage in communication. This is perhaps most prominently observed in the strong relationships between the frequencies of occurrence of various aspects of the language and their representation and processing by L1 speakers (Ellis, 2002; 2008). More specifically, the last 50 years of psycholinguistic research has showed that language processing is sensitive to usage frequency at all levels of language representation: phonology and phonotactics, reading, spelling, lexis, morphosyntax, MWS, sentence production, and comprehension (Ellis, 2002). In other words, implicit learning mechanisms enable language speakers to be sensitive to the frequency of occurrence across all linguistic levels (Ellis, 2002; 2008). Another framework closely associated with usage-based approaches to language acquisition is the construction-based approach in which an utterance is viewed as consisting of a set of constructions (e.g. Goldberg, 2006; O'Grady, 2005). Constructions are pairings between form and meaning including words and morphemes (e.g. *read -ing*,), MWS (e.g. *a cup of coffee*), and phrasal verbs (e.g. *get up*). The processing of language both in comprehension and production is fundamental to language acquisition. (Christiansen & Chater, 2016a). That is to say, each processing event creates an opportunity for language acquisition, and language acquisition in essence consists of learning how to process utterances (pp. 4-5).

Usage-based and construction-based approaches (Bybee, 1998; Christiansen & Chater, 2016a; Ellis, 2002; Goldberg, 2006; Kemmer & Barlow, 2000; Tomasello, 2003) have argued against a distinction between lexicon and grammar, or words and rules, or between stored and computed linguistic units. The rationale behind this argument is that language learning and processing are influenced by the amount of experience that language speakers have with the linguistic inputs. Thus all linguistic input, irrespective of its irregularity or idiosyncrasy, should be processed in a comparable way, and frequency effects should be observable in the processing of all types of linguistic input. As Bod (2006) notes, the allocation of representation to linguistic inputs is organised on the basis of statistics. In line with this, Bybee and McCllelland (2005) suggest that frequently used sequences become more accessible, and more entrenched. Therefore usage-based approaches to language acquisition generally view linguistic productivity as a gradually emerging process of storing and abstracting MWS (Christiansen & Chater, 2015, Tomasello, 2003). These approaches also enjoy mounting empirical support from psycholinguistic studies that language users process MWS quantitively faster than control phrases in both comprehension and production tasks (e.g. Arnon & Snider, 2010; Bannard & Matthews, 2008; Tremblay, et al. 2011; Jolsvai, McCauley & Christiansen, 2013; McDonald & Shillcock, 2003; Vilkaite, 2016) Therefore, usage-based approaches view MWS, alongside single words as building blocks of both written and spoken discourses.

Language presents different kinds of distributional information, including frequency, variability, and transitional probabilities (Erickson & Thiessen, 2015), and human mind is found to be sensitive to such statistics, this is known as implicit statistical learning (see section 1.7.1).

1.7.1 Implicit statistical learning.

There is increasing evidence that language users are sensitive to the statistical properties inherent in language (for an overview see Rebuschat & Williams, 2012). Along these lines, usage-based approaches put emphasis on statistical learning, typically viewed as a domain-general mechanism by which cognitive systems discover the underlying distributional properties of the input (Frost, Armstrong, Siegelman, & Christiansen, 2015). Specifically, language is an input rich environment with different types of distributional information including the frequency, variability and co-occurrence probability (Erickson & Thiessen, 2015). Studies on statistical learning has provided evidence that it is a mechanism responsible for phonological learning (e.g. Maye, Weiss, & Aslin, 2008), word segmentation (e.g. Saffran et al. 1996), syntactic learning (e.g. Thompson & Newport, 2007), and category formation (e.g. Gomez & Gerken, 2000). Furthermore, they provided evidence that both child and adult language learners can discover the underlying structural regularities of language by relying on distributional statistics (Frost et al. 2015; Saffran, Aslin, & Newport, 1996). Such statistical learning functions in both L1 (e.g. Saffran et al. 1996) and L2 acquisition (Frost et al. 2015). It is possible to say that there is a considerable evidence demonstrating the robustness of statistical learning, however, the nature of the cognitive mechanisms underlie statistical learning is still not yet fully understood. One conceptualisation of statistical learning is that unconscious statistical computations, also known as predictions, can be considered as learning cue-outcome relationships such as contingency learning (e.g. Misyak, Christiansen, & Tomlin, 2010; Williams 2009). An alternative conceptualisation is that speakers' sensitivity to frequencies and probabilities in language are needed for chunk formation rather than tracking predictive dependencies (Perruchet & Peereman, 2004).

In the next section (see 1.5.2), a review of the now-or-never bottleneck framework (Christiansen & Chater, 2016b) is presented. This framework is based on the speakers' basic information processing limitations. The now-or-never bottleneck is an integrated framework for language acquisition, processing, and evolution.

1.7.2 The Now-or-Never processing model.

Language processing takes place in the here-and-now; as we hear a sentence unfold, we quickly forget the preceding sentence. That is to say, if particular linguistic information is not processed immediately, that information is lost (Christiansen & Chater, 2016b). Furthermore, our memory for sequences of auditory and visual input is also very limited. For example, Miller's

(1956) found that our ability to recall auditory stimuli is far from being perfect. The imperfect memory capacity for both visual and auditory input combines with the fleeting nature of the linguistic input and they create a significant constraint on the language system. Therefore, if the linguistic input is not processed immediately, the new information will overwrite it (Christiansen & Chater, 2016b). At this point, how language comprehension is possible under these circumstances is a very important question. Christiansen and Chater (2016b) suggest that language users rapidly recode this input into chunks which are immediately passed to a higher level of linguistic representation. The chunks at this higher level are also passed to an even higher level of linguistic representation, leading to larger chunks of linguistic representation, this known as Chunk-and-Pass model (p. 5). During comprehension, our language systems use all available information to integrate the incoming information as quickly as possible to update the current information of what has been said so far. As the incoming acoustic information is chunked, it is immediately integrated with the contextual information to recognise the words. These words are then chunked into multi-word units. The recent psycholinguistic experiments showed that language speakers are sensitive to the MWS in on-line processing (e.g. Arnon & Snider, 2010; Wolter & Yamashita, 2017, Yi, 2018, see Chapter-5 for a detailed review of these articles). Similarly production requires running the process in reverse, starting with the intended message and gradually decoding it into increasingly more specific chunks (Christiansen & Chater, 2016).

To be able to chunk incoming information as rapid and as accurately as possible, our language systems rely on simple statistical information gleaned from sentences which provides powerful constraints on language comprehension (Christiansen & McDonald, 2009; Hale, 2006). Supporting this view, eye-tracking data suggest that speakers use a variety of sources of statistical information including phonological cues to syntactic contextual information to anticipate the processing of upcoming words (e.g. Altmann & Kamide, 1999; Staub & Clifton, 2006). Similarly, the results from event-related potential (ERP) experiments suggest that language users make quite specific predictions for upcoming words including lexical category, and grammatical gender (e.g. Hinojosa, Moreno, Casado, Munõz, & Pozo, 2005; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005). The predictive processing allows top-down information from higher-levels of linguistic representation to constrain the processing of the input at lower-levels. In order to facilitate chunking across multiple levels of representation, prediction takes place across the various levels of representation but at different time scales (Christiansen & Chater, 2016b p. 9). From the viewpoint of the now-or-never bottleneck, prediction allows the chunking of the linguistic input as early as possible.

Another important question for the now-or-never bottleneck framework is how new information can be acquired without interfering with prior information. If the learners have a global model of the entire language (e.g. a traditional grammar), they carry the risk of overfitting the model to capture the all the regularities in the momentary linguistic input at the expense of damaging the past linguistic input. In order to avoid this problem, Christiansen & Chater (2016b) suggest that learning local, consisting of learning about specific relationships between particular linguistic representation. That is to say, acquiring new items has implications for processing similar items, but learning current items does not bring major changes for the entire model of language. One way to learn in a local fashion is that learners store individual examples, and then those examples are abstractly recoded by successive chunk-and-Pass operations, and finally learners generalise piecemeal from those examples. An example of local learning is that MWS which were successfully chunked and processed before are processed faster than the control phrases (e.g. Arnon & Snider, 2010; Bannard & Matthews, 2008, Wolter & Yamshita, 2018). More generally, there is a processing advantage for the items which are based on the past traces of processing such as MWS. The principle of local learning was also respected by usage-based approaches (Bybee, 1998; Ellis, 2002; Goldberg, 2006; Kemmer & Barlow, 2000; Tomasello, 2003, Christiansen & Chater, 2016a). The principle of local learning is compatible with the frequency effects found at all levels of language processing (Bybee, 2007; Ellis, 2002; Tomasello, 2003). In brief, Christiansen & Chater, (2016b) argue that language processing is severely affected by limitations on human memory. To cope with this, our language systems rapidly chunk new materials at a range of increasingly abstract levels of representation. In the next section (1.6), a general discussion on the theoretical perspectives that attempt to explain the processing of MWS is provided.

1.8 A discussion on theoretical approaches

The key tenets of the some of the above theoretical positions are that MWS are stored and retrieved whole from memory rather than activating or accessing the individual words making up the MWS (e.g. Sinclair, 1991; Pawley & Syder, 1983; Wray, 2002). In other words, these positions imply that when language users encounter a high-frequency MWS such as a high-frequency collocation (e.g. *fast food*), they do not necessarily activate the individual components of this collocation. In order to investigate whether or not the individual words are accessed or play a role in the processing of MWS, researchers need to use a semantic or syntactic priming paradigm or explore whether individual word frequency affect the processing speed of the MWS (Siyanova-Chanturia, 2015). At this point, looking at the results of some of the psycholinguistic experiments which claim to have found evidence for holistic storage is an important step. In an eye-tracking study, Underwoord, Schmitt, and Galpin (2004) found that the terminal words in idiomatic expressions were processed faster than the same words in control phrases. Based on this finding, they concluded that the idioms were stored and accessed holistically. Furthermore, Jiang and Nekrasova (2007) conducted a grammatically judgment task to compare the speakers' the speed of reading for frequent phrases against the control phrases. The control phrases used in their study included ungrammatical sequences (e.g. *cars lots of a*). They found that the frequent phrases were read both more accurately and faster than the control phrases. They suggested that the MWS were recognised holistically and they were not subject to the full syntactic analysis. However, it is important bear in mind that the processing advantage does not shed much light on whether the MWS are stored holistically or not (Siyanova-Chanturia, 2015). The processing advantage only shows that MWS are processed quantitatively faster than control phrases used in these studies. Furthermore, none of these studies which reported to have found evidence for the holistic storage of MWS (e.g. Jiang & Nekrasova, 2007; Underwood et al. 2004) were not designed to explore the holistic storage because they did not investigate the activation of the individual words within MWS.

Compounds are in many ways very similar to MWS, in that a larger unit consists of smaller meaningful units (e.g. bedroom, blackboard). The question of whether or not individual parts are activated has been addressed in some of the compound processing studies. Badecker (2001) suggests that processing of compounds is purely compositional, whereas Juhasz (2007) argues that compounds are decomposed during recognition. More recent studies provided evidence that processing of compounds entails the activation of its individual components (Kuperman, Schrueder, Bertram, & Baayen, 2009; Mondini, Luzzatti, Saletta, Allamano, & Semenza, 2005). Overall, it is possible to say that individual constituents can be activated and accessed in the mental lexicon during compound processing. Furthermore, Arnon and Cohen Priva (2015) explored the effect of word and multi-word formation on the duration of words in a naturally elicited speech. More specifically, they looked at whether the relationship between word and multi-word information changes across the frequency continuum. They found that the effect of single word frequencies on phonetic duration decreased for highly frequent MWS, but they remained significant. However, the effect of the frequency of MWS on phonetic duration increased. Based on this finding, they suggested that the repeated usage leads to growing prominence of MWS, but the repeated usage does not eliminate the activation of single words This thesis therefore adopts the perspective that both the individual word and the phrase as whole affect the processing of MWS (see sections 5.1 and 5.2 for a more discussion on processing of MWS).

A processing advantage observed for MWS over novel phrases in comprehension (e.g. Arnon & Snider, 2010; Vilkaite, 2016; Wolter & Yamashita, 2017; Yi, 2018), and production tasks (e.g. Arnon & Cohen Priva, 2015; Bannard & Matthews, 2008; Janssen & Barber, 2012) is a vital piece of empirical evidence for the role of phrasal frequency in language processing and production. Furthermore, such a processing advantage demonstrates that language users are not only sensitive to the frequency of the single words but also to the phrases. This finding provides empirical support for usage-based and construction-based approaches to language acquisition (Bod; 2006; Bybee, 1998; Christiansen & Chater, 2016a; Ellis, 2002; Goldberg, 2006; Kemmer & Barlow, 2000; Tomasello, 2003). In these approaches, the allocation of representation to linguistic inputs is organised on the basis of frequency. The phrasal frequency effect of MWS reflects the language speakers' hundred and thousands of previous encounters with given MWS (Siyanova-Chanturia, 2015). Therefore knowledge of language should not be viewed as set of grammar rules, but as a statistical accumulation of linguistic experiences (Bod, 2006; Kemmer & Barlow, 2000). Each and every occurrence of a word or MWS contributes to its degree of entrenchment in the memory of the speakers. As language speakers' experiences with linguistics events change, the way they process and represent these events also change. For example, Wolter and Yamashita (2018) found that the more proficient L2 speakers are more sensitive to the frequency information encoded in the collocations, and thus their processing is more akin to the L1 speakers than the less proficient L2 speakers' processing.

The now-or-never bottleneck framework also provides important insights into the processing of MWS. For example, Christiansen & Chater (2016b) point out that prediction takes place at the various levels of representation to facilitate chunking across multiple levels of representation. One source of information that enables the prediction in the processing of MWS is the statistical knowledge implicit in language users' linguistic input, word-to-word contingency statistics or also known as transitional probabilities in psycholinguistic literature. The psycholinguistic experiments provided evidence that trasntional probabilities play a role in the processing of MWS (McDonald & Shillcock, 2003; Vilkaite, 2016; Yi, 2018). In other words, language users are likely to rely on the statistical information during the processing of MWS to anticipate the upcoming words. Furthermore, the local learning principle proposed by Christiansen and Chater (2016b), that is acquiring new items have implications for processing of similar items, but learning current items does not bring major changes for the entire model of language. This principle also brings some insights into the processing of MWS. There is a processing advantage for the items which are based on the past traces of previous processing experiences. The processing of the MWS is a good example for local learning principle. A number of studies have shown processing advantages for high-frequency MWS that were successfully chunked and processed by proficient L2 speakers (e.g. Sonbul, 2015; Wolter & Yamashita, 2018; Yi, 2018).

In the following chapter, methodological framework of the current study, combining corpus and experimental methods is presented. The current study including the target construction and research questions are also presented in the next chapter (see chapter 2).

Chapter 2: Methodological Approach of the Study

Language acquisition plays a central role in the study of human cognition and research on how we learn language cuts across the fields of developmental psychology, linguistics, education, and neuroscience (MacWhinney, 2017; Rebuschat, Meurers, & McEnery, 2017). It is a very challenging topic to study since both first and second language acquisition processes are highly complex. The complexity of the processes is increased further by the fact that both first and second language acquisition is heavily affected by individual learner differences (Rowland, 2013; Williams, 2012). Despite these and other important challenges, the researchers have made some significant progress in understanding how adult and children learn first and second languages (Rebuschat et al. 2017). The increasing range of methods and approaches played an important role in this conceptual and empirical progress (Hoff, 2011). The study of a complex phenomenon like language acquisition can significantly benefit from different methods and approaches (Rebuschat et al. 2017). Therefore, the current study uses a multimethod approach by combining corpus and experimental methods to study the processing of MWS. This chapter firstly discusses the benefits and challenges of combining corpus and experimental approaches to study language acquisition and processing and then it reviews the empirical studies that have combined corpus and experimental approaches (see 2.1). Secondly, some terminological clarifications are provided and finally the present study reported in this thesis is introduced, (see sections 2.2 and 2.3).

2.1 Converging corpus and experimental methods to study MWS

Psycholinguistics and corpus linguistics have had quite different research goals and approaches (Durrant & Siyanova-Chanturia, 2015). While corpus linguistics is mainly concerned with identifying patterns of use in large samples of language which aims to represent a particular speech community, psycholinguistics is concerned with the mental processes and representations of language during comprehension and production. Considering these different concerns of the two fields, unsurprisingly the attempts to combine them have been uncommon and difficult (Durrant & Siyanova-Chanturia, 2015; Gilquin & Gries, 2009). However, increasingly researchers from each field feel the need to cooperate with the other (see Rebuschat et al. 2017 for a discussion on combining corpus and experimental approaches). It should be noted that for a fruitful combination of corpus and experimental methods, a great care needs to be taken to consider the advantages and disadvantages of each methodology over the other, and in what ways they can complement each other. Gilquin and Gries (2009) argue that corpora involve data from natural contexts - thus they come with a higher degree of external validity than many experimental designs. Corpora also involve a larger range of data than that cannot be easily studied through experimental designs. For example, if 10,000 hits of a particular argument structure constructions are explored, the number of verbs in that construction is potentially considerably higher than they can be investigated in an experimental design. Experimental psycholinguistic methods also have many advantages. For example, they allow the exploration of language speakers' on-line cognitive processes. They make it possible to control the confounding variables. Furthermore, it is possible to investigate the constructions through experimental methods that are too infrequent in corpora given the Zipfian distribution. Since the advantages and disadvantages of corpora and experiments are largely complementary, using these two methods together is very helpful for (i) solving problems that would not be easily solved through single methods (ii) approaching phenomena from multiple perspectives.

Gilquin and Gries (2009) suggest that articles with a corpus linguistics focus that combine corpus and experimental methods are considerably rarer than the articles with psycholinguistics focus that combine the two methodologies. They suggest some ways in which the combination of corpus and experimental methods can be very helpful from a corpus linguistics perspective. Firstly, considering the fact that many corpora are very large, even the smallest findings might end up being significant. Therefore, additional experimental data will provide further insights that it will give the researchers a chance re-evaluate their findings and separate the wheat from the chaff. Secondly, different corpora might yield different results, and additional experimental evidence will help the researchers to have clearer understanding of the phenomena. Thirdly, corpus-based findings should be validated through experimental data. Finally, combining corpus and experimental data is helpful to gain insight into the relation between the two types of data (p. 17). The first two reasons are very straightforward and important, but the third reason may not seem very clear. For example, corpus linguistics have been developing various measures of association to measure collocational strength also known as transitional probabilities, however there is only little attempts to validate these measures through experimental methods, to demonstrate if any sensitivity L1 and L2 speakers would show for these measures (see section 3.1 for a detailed discussion on frequently used association measures). The final reason emphasises the extent to which corpus and experimental findings converge or diverge and researchers should offer explanations for the possible differences or similarities between two types of findings.

The combination of two methods can be very beneficial from a psycholinguistics perspective as well. Corpora are an indirect and approximate source of information about language users' experience with language use, which plays an important role in the mental processing and representation of linguistic input by L1 and L2 speakers (Rebuschat & Williams, 2012). Therefore, many studies used learner corpora to make psycholinguistic claims about second language learning. It should be noted that some topics of study is more suitable for this approach than some others. For example, as Gablasova at al. (2017a) suggest that corpora can provide direct information about the usage patterns of MWS produced by L1 and L2 users. In this regard, a few learner corpus studies have attempted to examine if adult L2 learners acquire and use language formulaically, that is relying on MWS rather than using the words-and-rules approach. Researchers have used learner corpora to examine if adult L2 learners' use of language lacks formulaicity (Durrant & Schmitt, 2010; Nesselhauf, 2005). One criticism of such attempts is that it is not possible to find out if a piece of language occurs in a corpus is or is not formulaic (MWS) for the language speaker who used it. Therefore, any direct inference made from corpus to mind needs further justifications (Durrant, 2013; Durrant & Siyanova-Chanturia, 2015), since the patterns found in corpora are the results of a variety of factors including psycholinguistic mechanisms and sociolinguistic conventions.

Recognising the fact that the productions of language appearing in corpora are the results of a range of factors that cannot be easily disentangled (Durrant, 2013; Durrant & Siyanova-Chanturia, 2015). The studies which base their psycholinguistic claims only on the corpus data tend to be criticised that their evidence undermines their conclusion (Durrant & Siyanova-Chanturia, 2015). Therefore, it is important for the researchers to complement their corpus finding with experimental data rather than making assumptions about psycholinguistic processes solely based on corpus data. The reason is that the patterns found in corpora cannot be solely based on psycholinguistic mechanisms, but they are also affected by sociolinguistic conventions. The first way of integrating the two types of methods is to use learner or general (reference) corpora as a source of experimental stimuli. The corpora are able to provide reliable quantitative information about language and researchers can use this information to create experimental stimuli (Durrant & Siyanova-Chanturia, 2015). This types of integration of corpus and experimental methods has been quite popular since psycholinguists are often concerned with exploring the effects of the types of variables which corpora can provide such as constructions' frequency of occurrence, distribution and levels of association between linguistic elements (see McDonald & Shillcock, 2003; Monaghan & Mattock, 2012). Using learner corpora in this way is not quite popular yet (see Millar, 2011; Siyanova & Schmitt, 2008 for examples). At this point, Gablasova et al. (2017) suggests that learner corpora are very useful resources to check whether low-frequency words or collocations are present in L2 speakers' lexicon (p. 172). This way learner corpora can also contribute to the stimuli preparation process alongside general corpora.

The second way of integrating corpus and experimental methods is to use corpora to confirm if the patterns which emerge from the psycholinguistic experiments are mirrored by the corpus data. (see Lowder & Godern 2016). The combination seen here reverses the common relationship between corpus and experimental methods seen in the first model. In the second model, the corpus is used to validate the findings emerging from the experiments. Since the vast majority of psycholinguistic experiments are lab-based and require participants to do the tasks with language that they do not normally do in real life situations, they are sometimes criticised for their limited external validity. Therefore, this way of integrating corpus and experimental methods to validate the findings emerging out of the experiments helps mitigating some of these criticisms. The third way of corpus and experimental methods involves drawing on data from both types of methods to triangulate on a single question (see Siyanova & Schmitt, 2008). In this model corpus is used both as a source of stimuli for the psycholinguistic experiment, and also used to triangulate the findings of the experiments. Next I review the empirical studies which have combined corpus and psycholinguistic methods in one of these three ways (see Table 2.1). To be able to review the literature of studies combining corpus and experimental methods I checked the databases, Web of knowledge, and PsychInfo. I used the search terms corpus linguistics, psycholinguistics, experimental methods, and multi-word sequences. To the best of my knowledge, the literature review below covers all of the empirical studies which combined corpus and experimental methods to investigate MWS and similar constructions such as relative clauses.

Table 2.1 Studies combining corpus and experimental methods

Study	Target language and Structure	Patricipants	Design
Siyanova & Schmitt (2008) explored the use and mental processing of collocations by L1 and L2 speakers of English.	English adjacetive-noun collocations (e.g. *strong tea*)	60 L1, 60 L2 speakers of English	First, they carried out a corpus study to explore use of collocations by frequency and association. Second, they explored L1 speakers' and learners' processing of collocations using an on-line and off-line acceptability judgment tasks.
Wulff (2009) explored the variables affecting the idiomaticity of the V NP constructions.	English V NP idiomatic constructions (e.g. *take the piss*)	39 L1 speakers of English	Wulff analysed the compositionality, syntactic, lexico-sytactic and morphological flexibility of the constructions (BNC). Then, participants were asked to rate the idiomaticity of the constructions in an off-line task.
Ellis & Simpson-Vlach (2009) investigated how corpus-based measures of formulaicity affect both instructional validity and mental processing of MWS	The MWS that are commonly used in academic contexts (e.g. *in other words*)	12 L1 speakers of English participants took part in the processing and production tasks	They used the MICASE and selected academic BNC files to identify target structures. Then, they investigated L1 speakers' mental processing /productions of those MWS.
Millar (2011) investigated the L1 English speakers' processing of collocations that deviate from the target language norms.	Adjective-noun collocations produced by learners that deviate from the target language norms (e.g. *best partner*, instread of *ideal partner*)	30 L1speakers of English	A corpus of English learner essays was used to identify collocations that deviate from the target language norms. Both learner and correct versions (BNC) were presented in a self-paced reading experiment.
Monaghan & Matttock (2012) investigated the interactions between word co-occurrence constraints and cross-situational statistics in word learning.	Learning referring and non-referring words.	45 L1 speakers of English	They first conducted a corpus analysis of child-directed speech. Then they used this information to construct an artificial language that is based on natural language statistics.
Reali (2014) investigated the processing of relative clauses in Spanish by using corpus and experimental methods.	Spanish relative caluses	30 L1 speakers of Spanish	She first conducted corpus analysis of relative clauses' distrbutional patterns. She then carried out a series of experiments to explore if the findings mirror the corpus statistcs.

Lowder & Gordon (2016) investigates the effects of sentence structure on the processing of complement coercieon.	Complement coercion (e.g. began the memo)	36 L1 speakers of English	First they designed an eye-movement experiment to examine the effects of sentence structure on the processing of complement coercieon. Then, they conducted a corpus study to investigate the extent to which patterns found in the experiments mirrored in naturally occurring language.

2.1.1 The review of studies combining corpus and experimental data.

Siyanova and Schmitt (2008) explored advanced level L2 learners' knowledge of collocations using learner corpus, and off-line and on-line judgment tasks. In the learner corpus study, they extracted a total 810 adjective-noun combinations from a corpus of thirty-one essays, part of the International Corpus of Learner English. They found that nearly half of the collocations that occurred in the learner corpora appeared in the BNC at least six times. Nearly a quarter of them were not attested in the BNC. To find out whether advanced learners process collocations in a native-like way, L1 speakers and advanced level L2 English learners were asked to judge the acceptability of the collocations in on-line and off-line acceptability judgment tasks (AJTs). One merit of the study reported in Siyanova and Schmitt (2008) was that it investigated the learners' knowledge of collocations from various perspectives using a multimethod approach to have an in-depth analysis of the phenomenon. It is particularly important that Siyanova and Schmitt (2008) discussed the differences between learner corpus and on-line processing studies in terms of demonstrating the differences between learners' and L1 speakers' collocational knowledge. The main findings of the study were that the majority of the collocations produced by the learners were appropriate. However, the AJT revealed that learners had considerably poorer intuitions of the collocations than L1 speakers. L1 speakers reliably distinguished low- medium- and high-frequency collocations but L2 speakers were only able to distinguish low-frequency collocations from the high-frequency ones.

Wulff (2009) combines experimental and corpus methods to investigate the idiomatic variation from a usage-based approach to language acquisition. She also explored which features are responsible for the speakers' perceptions that constructions are idiomatic and how these features can be defined from a corpus linguistics perspective. She used the British National Corpus (BNC) to extract the V NP idioms, and the corpus data were complemented with overall idiomaticity judgments for the extracted 39 idioms from the BNC in an off-line task. From a usage-based perspective, she developed a corpus-based compositionality and flexibility statistics. She demonstrated that how different parameters that correlate with idiomaticity relate to one another and whether they can predict participants' judgment of idiomaticity. She found that L1 speakers of English relied on multiple factors simulatanously. Lexico-syntactic and morphological flexibility played greater roles than the other factors. Wulff (2009) argued that combining the experimental and corpus methods enabled this study to adopt a multifactorial approach that weighs the relative influence of each variation parameter (e.g. flexibility). Moreover, this approach made it possible to objectively define the parameters such as different types of compositionality (p. 147).

Furthermore, Ellis and Simpson-Vlach (2009) investigated how corpus-based measures of formulaicity affect the mental processing of MWS. They explored L1 speaker participants' on-line processing of MWS which are specific to academic speech and writing. They extracted three, four, and five-word MWS occurring at least 10 times per million words from various corpora of academic language such as MICASE (Simpson et al. 2002) and selected academic BNC files. To measure the associations of MWS, they used mutual information (MI). They also extracted the raw phrasal frequency scores of each extracted MWS (see section 3.1 for a detailed discussion on association measures). They chose a sample of 108 MWS, 54 from speech and 54 from written genres to investigate the processing of these MWS in series of comprehension and productions tasks. Ellis and Simpson-Vlach (2009) suggested that the combination of corpus and psycholinguistics experiments showed that the statistics of language corpora, particularly the MI-score, is implicitly represented in language users. They found that phrasal frequency and transitional probabilities (as measured by MI-scores) predict processing of MWS for L1 spakers.

Millar (2011) investigated the effects' of L2 learners' use of atypical (non-nativelike) collocations on L1 speakers' processing. Millar (2011) is one of the rare studies which relied on learner corpora to prepare experimental stimuli for a psycholinguistic experiment. He firstly identified atypical collocations from a 180000 word corpus of essays written by 96 L1 Japanese-English L2 learners. His criteria for considering word pairs as atypical collocations were that they should occur at least twice, appear in the writing of at least two learners and they should not occur in the BNC. After that he identified the typical collocations which convey the learners' intended meaning. For example, the atypical word pair *marriage life* was matched with typical collocation *married life*. He completed this process by searching for the atypical pairs components words in the BNC. The pairs were put in appropriate sentence contexts. L1 speaker participants were shown the sentences in a self-paced reading experiment. He found that learners' deviations from the target language collocations creates an increased processing burden for L1 speakers. Millar's (2011) is a great example of how both general reference corpora such as the BNC and learner corpora can contribute to the preparing material for psycholinguistic experiments to explore the processing of MWS.

The research reported by Monaghan and Mattock (2012) provided a great example for how corpus and experimental work complement each other. They investigated the interactions between word co-occurrence constraints and cross-situational statistics in word learning. Their first study was a corpus analysis of child-directed speech addressed to six children. They specifically looked at the utterances that contained at least one noun. Then they counted the number of words from other categories for each sentence containing at least one noun. They found that the typical exposure the child experiences is either just one word as a potential referent or no nouns at all. The utterances with one or more nouns also include verbs, adjectives, adverbs and function words. The corpus analysis also showed that children are exposed to utterances involving several nouns potentially referring to the objects in the environment. However, more frequently children are exposed to the utterances that contain several words other than nouns. Then they used this information derived from the corpus analysis to build an artificial language that is based on the statistics of child-directed speech directed to six children. They found that word learning situations present ambiguity because of the range of possible referrents for each words and multiple words in each situation - only some of which map onto the objects in the environment. Considering the fact that artificial language research is sometimes criticized for its limited ecological validity, the use of distributional statistics from corpora may increase the ecological validity (see also Monaghan & Rowland, 2017).

Another impressive example of multimethod research involving corpus and experimental work was Reali (2014). She combined corpus and psycholinguistic experimental methods to investigate whether levels of entrenchment suggested by corpus data produce a measurable effect for Spanish L1 speakers' processing of relative clauses. She firstly conducted a corpus analysis to investigate distributional patterns of Spanish relative clauses. More specifically, she looked at whether there is a preferred word order for object and subject relative clauses by L1 speakers of Spanish. She used a free access corpus of spoken Spanish that included a variety of conversational topics. She extracted a total of 1913 relative clauses of which 1409 were subject relative clauses, and 564 object relative clauses. She found that object relative clauses have higher levels of surface structure flexibility than subject relative clauses. She then carried out a series of psycholinguistic experiments (complexity ratings of relative clauses and self-paced reading) and found that processing of the nested structures such as relative cluases is affected by the experience-based factors. These findigs mirror the corpus statistics. They further suggest that the representation and processing of nested constituents may be facilitated by exposure and usage of frequent word chunks. The advantage of combining corpus and experiments for this study was that the corpus data revealed a fine-grained frequency patterns of relative clauses, and these patterns set the starting point to test the predictions of usage-based approaches to language acquisition.

The research described in Lowder & Gordon (2016) was an interesting example of multimethod research combining corpus and experimental methods. Differently from the studies reported above, Lowder & Gordon (2016) firstly carried out two eye-tracking experiments to investigate the effects of sentence structure on the processing of complement coercion, and then they conducted a corpus study to examine the extent to which patterns emerged from the experiments were mirrored in naturally occurring language. The experiments showed that the processing difficulty associated with complement coercion is reduced when the event selecting verb and entity denoting noun appear in separate clauses compared to when they appear in the same clause. Using the Corpus of Contemporary American English, they conducted a corpus analysis of event-selecting verbs embedded in the relative clauses. They firstly randomly selected 1000 instances of the verbs they used in the experiments (i.e. *attempted*, *began*, *endured*, *finished*, *mastered*, *resisted*, *started*, *tried*). Then the selected sentences were coded as to whether the target verb was part of an object or subject relative clauses. The results of the corpus study and the experiments showed that processing cost is associated with complement coercion is reduced when the event-selecting verb and entity denoting noun phrases appear in separate clauses.

To sum up, as can be seen in the examples above, corpora can be used to prepare stimuli for psycholinguistic experiments (see Ellis & Simpson Vlach, 2009; Monaghan & Mattock, 2012). Corpus studies can also help psycholinguists to generate hypotheses that they can test in experimental designs (see Reali, 2014; Siyanova & Schmitt, 2008). Furthermore, corpus studies can also be used to validate the findings of the psycholinguistic experiments (see Lowder & Gordon, 2016). While the combination of corpus and psycholinguistic methods has a lot to offer, researchers need to pay attention to many details. One important problem of interdisciplinary studies is that misappropriation or misguided use of methods that have developed within a particular discipline (Durrant & Siyanova-Chanturia, 2015). This particularly involves methodological choices related to the use of statistics since corpus linguists might tend to conflate data from many speakers, whereas psycholinguists might want to measure individual differences of processing of one type of construction (see Durrant & Schmitt, 2009; Gablasova Brezina, McEnery, 2017b for quantifying data each individual separately). More generally, to avoid this type of problems both corpus linguists and psycholinguists need to familiarise themselves thoroughly with the methodologies they want to adopt (Durrant & Siyanova-Chanturia, 2015). As future directions for the combination of corpus and experimental approaches, Monaghan and Rowland (2017) suggest that accessibility of large corpora for exploring both first and second language acquisition is needed. The corpora should be collected in a way that allows observing individual differences between learners, and that makes it possible to conduct longitudinal analyses; this would be very helpful to complement experimental works (p. 29).

2.2 Terminological clarification and the target construction

The study of MWS has cut across the fields of psycholinguistics, corpus linguistics, and language education, and the many terms have been used in research around the formulaicity of language. The variation in terminology also reflects the differences of researchers' focus and different aspects of the phenomena investigated (Myles & Cordier, 2016). For example, the term *chunk* is commonly used in psycholinguistic research, whereas the term *clusters* is widely preferred in corpus linguistics (Myles & Cordier, 2016). It is also more problematic to see that the same term may be used by different researchers to refer to constructs that they might overlap but are essentially different. The term *formulaic sequences* coined by Wray (2002), has been widely adopted and used by various researchers. On the one hand, some researchers adopted term *formulaic sequences* as an umbrella term (Weinert, 2010; Wood, 2015). On the other, some researchers used the term to refer to the idioms and idiomatic expressions that are assumed to be processed holistically (e.g. Underwood, Schmitt & Galpin, 2004). As Wray (2012) also noted, this terminological confusion is potentially problematic, especially when some claims are made about formulaic sequences in general, but the focus of the investigation was only one type of formulaic sequence. To avoid any terminological confusion, I used multi-word sequences (MWS) and chunks as an umbrella terms, which were also used by various studies with psycholinguistic focus (McCauley & Christiansen, 2017; Christiansen & Arnon, 2017). However, the current study specifically focuses on collocations and the findings of both the corpus study (see chapter 4), and the experimental study (see chapter 6) concern the collocations rather than the other type of MWS. The next section introduces the target construction of this study (see 2.2.1).

2.2.1 Target construction.

Collocations are a prominent type of MWS, that have received a special attention in both corpus-based language learning and psycholinguistic studies in the last decade. This study looks at the processing of high- and low-frequency collocations in Turkish and English, and by L1 and L2 speakers. One of the main reasons why collocations have received a particular attention is that they are lexical patterns that are shaped by more conventions within the language rather than grammatical or semantic restrictions (Wolter & Yamashita, 2015). For example, it would be natural to say *strong tea,* but any experienced speaker would notice the comparative novelty of *dark tea* in English. However, in Turkish for example, the exact opposite would be true. Another reason why they have received special attention is that efficient language processing and use are, to an important extent, contingent upon the formation of systematic and meaningful links between words in the lexicon, and the knowledge of collocations help L2 learners to develop an efficient lexical network (see part 2 chapter for a review of collocational properties). In addition, collocational knowledge therefore is considered as an important component of one's overall language competence, by many researchers (e.g. Bestgen & Granger, 2014; Ellis et al. 2008; González Fernandez & Schmitt, 2015; Hoey, 2005; Pawley & Syder, 1983; Sinclair, 1991).

Different approaches to operationalising the MWS, specifically the collocations have been noted in the literature (McEnery & Hardie, 2011, p. 122-133). The two most widely known approaches are the "phraseological", and "frequency-based" approaches. The phraseological approach focuses on establishing the semantic relationship between two or more words, and the degree of non-compositionality of their meaning (Nesselhauf, 2004, Howarth, 1998). In this approach, the collocations are not simply free combinations of semantically transparent words, but they follow selectional restrictions (e.g. *'slash' one's wrist* rather than *'cut' one's wrist*). The empirical or frequency-based approach draws on quantitative evidence of word co-occurrence in corpora (Evert, 2008; Gablasova et al. 2017; Hoey, 1991; Gries, 2013; Sinclair, 1991) In traditional corpus linguistics, collocations have been described as the relationship between two words which occur near each other (see Sinclair 1991 for example). As the development of new generation corpus tools (e.g CQPweb, #LancsBox), this approach involved more sophisticated statistical measures, which are known as association measures (AMs), to identify the psychological association between words, which is evidenced by their co-occurrence in corpora (see section 3.1 for a detailed discussion about association measures). Not surprisingly, there are advantages and disadvantages to both approaches. It is noteworthy that in the phraseological approach, the operationalisation of collocations - whether they are free or restricted combinations of words - is quite problematic because the criteria rely heavily on intuition. However, semantic transparency is a variable, which is likely to affect the processing of collocations. We cannot expect that free combinations such as *pay a bill*, which are used in literal meanings would be processed in exactly the same way as more restricted word combinations such as *pay a visit,* in which one of the components are used in a figurative sense. Therefore, it would be very interesting to investigate the effects of semantic transparency on collocational processing. For the present study (see section 2.3), however, it was necessary to adopt an approach that was rooted in frequency-based tradition, because this study is centrally concerned with the extent to which collocational frequency and single word frequency counts affect the mental processing of collocations in typologically different languages: English and Turkish and for L1 and L2 speakers.

2.3 This dissertation

The study of MWS, also known as the formulaic perspective, has provided important insights in a range of fields including corpus linguistics, psycholinguistics, language pedagogy, and theoretical linguistics (see also Wray, 2008; Durrant, 2013; Gablasova et al. 2017a; Siyanova-Chanturia & Martinez, 2014). However, as Durrant (2013) also suggests, an important weakness in this field of study is that it focusses only on narrow range of languages, particularly on English. That is to say, the benefits of adopting a formulaic perspective was limited to these narrow range of languages, and also the status of MWS as a general feature of language remains insufficiently established. It is therefore important that the study on MWS should focus on languages with different typological profiles. At this point, Biber (2009) noted that English has minimal inflectional morphology but a large set of grammatical function words. These two factors are the central determinants of MWS that are common in English. However, there is only little research on MWS realised in morphologically rich agglutinating languages such as Turkish and Finnish. The agglutinating languages are interesting areas of exploration for formulaic perspectives since they employ extensive system of suffixes to build up complex word forms. In other words, in these languages, formulaic patterns are found within words (see Durrant, 2013), as well as between words. Put simply, some morphemes co-occur more frequently than others. In this regard, Durrant (2013) suggests that verbal, inflectional, high-frequency morphemes enter into collocational relations with their syntagmatic neighbours. This study is primarily interested in the extent to which the same mechanisms such as single-word and phrasal frequency affect the processing of MWS in typologically different languages, English and Turkish. Furthermore, it also investigates to what extent the same mechanisms affect the processing of collocations in L1 and L2. The next section provides some background information about the morphological structure of Turkish, an agglutinating language.

2.3.1 Turkish, an agglutinating language.

Turkish, like Hungarian and Finnish, is an agglutinating language, including words that are very long and complex. The suffixation is the main word formation process, in which the new words are formed by attaching an affix to the right of a base form, also known as the root. Traditionally the Turkish suffixes are divided into derivational and inflectional suffixes. Derivation is defined as the creation of a new lexical item (Göksel & Kerslake, 2005 p. 52). Inflectional suffixes, on the other hand, are used to mark functional relations such as case, person, and tense Göksel & Kerslake, 2005 p. 68). The suffixes that can be attached to nominals are known as nominal inflectional suffixes, and they are used to mark number, possession and case. Turkish has five case suffixes: accusative, dative, genitive, locative, and ablative. Unlike derivational suffixes, inflectional suffixes are perceived as productive (see examples 1a to 1e). There are also examples of noun inflections in Turkish where plural, person, and case occur together (see examples 2a to 2b).

(1a) suyu (the water)
water-ACC

(1b) suya (to the water)
water-DAT

(1c) suda (in the water)
water-LOC

(1d) sudan (from water)
water-ABL

(1e) sunun (of the water)
water-GEN

(2a) sularının (of the waters)
water-PL-P3S-GEN

(2b) sularında (in their waters)
water-PL-P3S-LOC

2.3.3 Research questions.

The present study sets out to investigate the factors affecting the processing of two-word adjective-noun collocations in Turkish and English. Before setting up a psycholinguistic experiment to look at the online processing of the collocations in the two languages, this study firstly designed a corpus study to contrastively explore collocations' frequency of occurrence and association statistics in English and Turkish. The main goal of this corpus study is to explore in what ways language typology, particularly the agglutinating structure of the Turkish language affects the collocability of adjectives and nouns. In this corpus study, frequency counts and association statistics of Turkish and English adjective-noun collocations are contrastively explored with the aim of addressing two research questions below (see chapter 4 for how these research questions are addressed).

i. How different (or similar) are frequency counts and association statistics for translation-equivalent English and Turkish adjective-noun collocations?

As Durrant (2013) also suggested, word combinations in Turkish are predicted to be less frequent than their equivalents in English. Since meanings which require multiple word expressions in English can be expressed using a single word in Turkish, individual word forms (lexemes) are expected to have lower frequency counts than their English equivalents. Therefore, collocations and other type of MWS are also predicted to have, on average, lower frequency scores in Turkish than their equivalents in English.

ii. How different (or similar) are the frequency and association scales for unlemmatised and lemmatised English and Turkish adjective-noun collocations?

Due to the agglutinating structure of Turkish, I would predict to find a notably larger difference between the unlemmatised and lemmatised Turkish collocations' frequency and association counts than English collocations' frequency and association counts.

The psycholinguistic experiments set out to ascertain the extent to which L1 English and Turkish participants rely on the same mechanisms for processing adjective-noun collocations. Furthermore, the present study explores whether the participants' response times (RTs) for adjective-noun collocations mirror the patterns emerged from the corpus study, in relation to the collocations' frequency of occurrence and collocational strength. Moreover, the present study ascertains whether the same factors affect L1 English and advanced L1 Turkish-English L2 learners' processing of collocations (see chapter 6 chapter for how these research questions are addressed).

i. Do both L1 speakers of Turkish and English process the high-frequency collocations faster than the low-frequency collocations, and the low frequency collocations faster than the baseline items?

L1 speakers of both Turkish and English will process the high-frequency collocations faster than the low-frequency collocations, which will in turn would be processed faster than the baseline items. This prediction is largely based on the findings of the previous studies that both L1 and advanced L2 speakers are sensitive to the frequency of the MWS (e.g. Arnon & Snider, 2010; Wolter & Yamashita, 2018; Yi, 2018)

ii. Does L1 speakers of Turkish process both high- and low-frequency collocations more rapidly than L1 speakers of English?

L1 speakers of Turkish will process both high-frequency and low-frequency adjective-noun collocations more rapidly than L1 speakers of English. This hypothesis is based on findings of the corpus study (see Chapter 4 for the corpus study) that lemmatised collocations have considerably higher collocational strength in Turkish than their equivalents in English.

iii. Do lemmatised collocation level frequency counts have a larger effect on the processing of Turkish collocations than English collocations?

Lemmatised collocation-level frequency counts will have a larger effect on the processing of Turkish adjective-noun collocations than English adjective-noun collocations. This hypothesis is based on results of a corpus study (see Chapter 4 for the corpus study).

iv. Do single word-level frequency counts have a larger effect on the processing of Turkish collocations than English collocations?

Word-level frequency counts of the nouns will have a larger effect on the processing of Turkish adjective-noun collocations than English adjective-noun collocations. This hypothesis is based on the results of the corpus study (see Chapter 4 for the corpus study) that Turkish adjective-noun collocations have many more inflected forms than their equivalents English.

v. Do both L1 English and advanced level L1 Turkish-English L2 speakers show sensitivity to both collocational and single-word level frequency counts?

L1 Turkish-English L2 speakers (advanced) will show sensitivity to both collocational and single-word level frequency counts. This hypothesis is based on the result of previous psycholinguistic studies (e.g. Arnon & Cohen Priva, 2015; Wolter & Yasmashita 2018), that the L1 and proficient L2 speakers attend to both single word and collocational frequency counts simulatanously.

Part 2: A Corpus Study of Adjective-noun Collocations in Typologically Different Languages: English and Turkish

It has long been noted that multi-word sequences (MWS) are a pervasive feature of language (see Biber 2009; Brezina, McEnery, Wattam 2015; Durrant & Schmitt, 2010 for corpus linguistic perspectives; Siyanova-Chanturia & Martinez, 2015; Wolter & Yamashita, 2018 for psycholinguistic perspectives). The corpus-based research on MWS dates back to earliest studies of collocations. Firth's (1957 p. 179) suggestion that "you shall know a word by the company it keeps" has been one of the most cited quotes in corpus linguistics. Unsurprisingly, MWS have been the focus of interest in corpus linguistics, because corpus is able to provide a rich source of information about the regularity, frequency, and distribution of MWS in languages (Gablasova et al. 2017). More specifically, corpus linguistics research has offered techniques to identify the MWS in written, and spoken discourses, and it has contributed to the definitions of the MWS as they were documented in corpora (Evert, 2005; Gries, 2008). Therefore, this part approaches MWS through the lens of corpus linguistics. It firstly provides a review of previous studies (see Chapter 3) and secondly reports on a comparative corpus study of adjective-noun collocations in English and Turkish (see Chapter 4). Chapter 3 firstly introduces the collocational properties and then it provides a critical review of corpus-derived AMs (see section 3.1). It secondly reviews the corpus-based contrastive studies on MWS (see section 3.2). Finally, it reviews corpus studies focussing on formulaicity in agglutinating languages (see section 3.3). Chapter 4 analyses frequency and collocational strength of adjective-noun collocations in English and Turkish from a contrastive perspective, using the British National Corpus (BNC) and Turkish National Corpus (TNC), that are balanced general reference corpora of the respective languages. Furthermore, chapter 4 includes methodology (see 4.1), results (see 4.2), discussion (see 4.3).

Chapter 3: Literature Review of Corpus Studies on MWS

Collocations as a particular type of MWS has received special attention in corpus linguistics research in the last 10 years and this attention has been even growing recently (Durrant & Schmitt, 2009; Gablasova et al. 2017; González Fernández & Schmitt, 2015; Nguyen & Webb, 2016). Since collocations are lexical patterns that are shaped by conventions within the language rather than grammatical or semantic restrictions, they offer a way of describing linguistic knowledge with respect to language use. In this regard corpora as a large databases of language usage data provide evidence of language users' co-selection of words, which reveal regularities in the use of collocations. Relying on corpora, researchers are able to hypothesise the factors affecting, the use, acquisition and mental processing of collocations. To be able to understand the phenomena of collocations, we, the researchers need to be aware of their properties. Therefore, the section 3.1 briefly discusses the collocational properties highlighted by the frequency-based approach. Another crucial aspect of the corpus-based research on collocations is that selection and interpreting the findings based on assosication measures (AMs). Since they directly and significantly affect the findings of the studies and consequently the insights into language learning and processing they provide, they are integral part of the research examining the use or mental processing of collocations. Therefore, this study pays a special attention to AMs (see also section 3.1).

As previously discussed in the introduction section, the corpus studies showed that a high proportion (up to 50%) of the language produced by adult L1 speakers in both spoken and written discourses consists of MWS (Erman and Warren, 2000; DeCock et al. 1998; Howarth, 1998). Furthermore, Biber et al. (1999), looking at written and spoken corpora separately, reported that MWS constituted 28% of the spoken, and 20% of the written discourse analysed.

Although they are typically found to comprise a high proportion of the written and spoken texts, surprisingly there is only a few corpus studies focusing on various aspects of MWS in different languages contrastively (e.g. Cortes, 2008; Granger, 2014). In line with this, Durrant (2013) notes that research in formulaicity focusses only on narrow range of languages particularly on English. To be able to understand the roles of MWS in language processing and use, we need more data from different languages, particularly languages with typologically different profiles. In this regard, this study reviews the previous corpus-based contrastive studies examining the MWS in different languages (see section 3.2), and then, it touches on the MWS in agglutinating languages (see section 3.3). The literature reviews presented in sections 3.2 and 3.3 also enables us to clearly observe the research gaps that the corpus study reported in chapter 4 addresses.

3.1 Collocational properties and selecting and interpreting association measures

The frequency-based approach highlighted the following properties of collocations: distance; frequency; exclusivity; directionality; dispersion; type-token distribution; and connectivity (Brezina et al. 2015; Gries 2013). The distance specifies the span around a node word where we search for collocates. The span also known as collocational window. Depending on the researchers' interest, the collocation window can be as small as one word or as large as possible. The frequency of use highlights how many times a collocation occurs in corpora. This is also known as collocational frequency. For instance, the adjective-noun collocation *good job* appears to occur frequently in a general corpus of English, the British National Corpus (BNC), given its normalised frequency count per million words of 7.92. However, it is important bear in mind that the adjective *good* is a very frequent adjective (normalised frequency count per million words of 722.08 in the BNC), and it modifies many other nouns such as *good reason*, *good news*, and *good morning*. Therefore, the relationship between adjective *good* and noun *job* cannot be described as a very exclusive one. The exclusivity highlights the high possibility that when a word appears in the text, the following word would be predicted by the readers. For instance, the adjective *part-time* is fairly exclusively connected with the noun *job* (see Brezina, et al., 2015, for a more detailed discussion on distance; frequency; exclusivity).

Directionality highlights the fact that the strength of association between two words is predominantly asymmetrical. For instance, the adjective *part-time* has a stronger connection with the noun *job*, than the noun *job* with the adjective *part-time* in the BNC because the word *job* co-occurs with other words more frequently than the word *part-time* does (see Gries, 2013 for a detailed discussion directionality). Dispersion refers to the property of how widely the collocation distributed across the texts in the corpus. For example, the adjective *special* collocates with the noun *time* in 63 cases (raw frequency count) and distributed across 57 texts. This is a quite evenly distributed collocation compared to another potential adjective collocate of noun *time*. Adjective *imaginary* also collocates with noun *time* in 63 cases, but they only appear in 4 texts in the BNC (see Gries, 2008 for a detailed discussion on dispersion). The type-token distribution takes into account the level of competition for the slots around the node word from other collocates (see Gries, 2013, for a more detailed discussion on type-token distribution). To these properties, Brezina, et al. (2015) added a seventh property, which is the connectivity between individual collocates. Connectivity highlights the fact that collocates of words are part of a complex network of semantic relationships. For instance, the adjective *part-time* does not collocate with the adjectives *wonderful*, *decent*, *well-paid* in the BNC, but they are all connected through the noun *job*. In other words, the noun *job* collocates with all four of these adjectives, *part-time, wonderful*, *decent*, and *well-paid*.

In order to identify collocations based on these seven properties, it is crucial for corpus linguists to select, use, and interpret findings based on various association measures (AMs). Corpus-based language learning and psycholinguistic studies using statistical definitons of collocations have distinguished between two major features of collocability (Ellis et al. 2015), absolute frequency, and strength of association between word pairs. While the frequency only approach is solely based on counting the co-occurrences of word pairs, AMs combines information about frequency with other collocational properties that can be expressed mathematically: dispersion, exclusivity, and direction (Evert, 2008, Gablasova et al. 2017; McEnery & Hardie, 2011). However, AMs are often grouped together and unhelpfully labelled as measure of collocational strength. If possible, they should be kept separate because they refer to the different properties of collocations and they are likely to play different roles in the processing of collocations (Gablasova et al. 2017). Although there are over a dozens of available AMs (see Evert, 2005; Pecina, 2010; Wiechman, 2008 for detailed review of AMs), so far only a few AMs such as MI-score, and t-score have been used in Second Language Acquisition (SLA) and psycholinguistic research predominantly (e.g. Bestgen & Granger, 2014; Durrant & Schmitt, 2009; Ellis, Simpson-Vlach, & Maynard, 2008; Williams, 1998). Although some AMs (e.g. t-score, MI-score) have been largely used in research, often their mathematical properties do not appear to have been fully explored (Gablasova et al. 2017). It is therefore important to discuss a few largely used AMs, with regard to their mathematical reasoning, the scale on which they operate, and the properties of the collocations they highlight. Furthermore, the discussions of AMs in this part will shed light on the choosing the right AMs for the contrastive study of collocations in English and Turkish (see Chapter 4). After reviewing the largely used AMs (e.g. MI), a recently introduced measure of association Log Dice is also reviewed.

The t-score has been defined as a measure of certainty of collocations (Hunston, 2002, p. 73; Bestgen & Granger, 2014), and of the “strength of co-occurrence”, which is used to test null hypothesis (Wolter & Gyllstad, 2011). Despite its large use in SLA research, it is important to have a look at arguments raised against these definitions of t-score, and its misguided use to extract collocations. Gablasova et al. (2017) argued that these definitions are not “particularly accurate”, because the t-score does not have a clear mathematical grounding. Furthermore, Evert (2005) suggests that t-score cannot be used to reliably establish the rejection region of null hypothesis. At this point, it is important to look at how the t-score is calculated. It is the subtraction of random co-occurrence frequency from the raw frequency. Then this is divided by the square root of the raw frequency. Since raw frequency scores directly depend on the corpus size, t-score directly affected by the size of corpora used. Therefore, it clearly does not operate on a standardised scale, and are not “comparable across corpora of different sizes” (see Gablasova et al. 2017 p. 8-9, Evert, 2005 p. 82-83 for a more detailed discussion of t-score). In practice, the t-score is used to highlight frequent combinations of words (e.g. Durrant & Schmitt, 2009; Siyanova & Schmitt, 2008). For example, Wolter and Gyllstad (2011) used t-score of 2 as a threshold for identifying frequent collocations. It is stressed in the literature that there is a strong correlation between the t-score and raw frequency (Durrant & Schmitt, 2009). Gablasova et al. (2017) suggests that for the top 100 t-score ordered bigrams in the BNC, the t-score strongly correlates with their raw frequency (r = 0.7); however, the correlation is considerably weaker (r = 0.2) in the top 10000 bigrams. Thus, the t-score and the raw frequency should not be seen as co-existential terms as suggested in the literature.

The MI-score has been largely used in corpus-based language learning and psycholinguistic research with various labels; such as a measure of strength (Hunston 2002), tightness (González Fernández, Schmitt, 2009), coherence (Ellis, Simpson-Vlach, & Maynard, 2008), and appropriateness of collocations (Siyanova & Schmitt, 2008). The MI-score is based on a logarithmic scale to express the ratio between the frequency of the collocation, and the frequency of the random co-occurrence of the two words in the combination (Church & Hanks, 1990 p. 23). A MI value of 0 indicates that a word combination co-occurs as frequently as expected by chance. A negative MI value shows that a word combination co-occurs less frequently than expected by chance (Evert, 2008). Although the MI-score is based on a normalised scale that is comparable across corpora (Hunston, 2002), it does not have a theoretical minimum and maximum values since it is not scaled to a specific range of values. (Gablasova et al. 2017). In addition, the MI-score has been found to favour low-frequency combinations, particularly the word combinations from large corpora (Evert, 2008 p. 1226, Gablasova et al. 2017 p. 10). Therefore, we should be careful not to automatically interpret the larger values, as an indication of stronger collocations. The bigger the MI-score is, the more exclusively associated and the less frequent the collocation is (see Gablasova et al. 2017 p. 18-19 for a more detailed discussion). Many psycholinguistic and CL studies used MI-score of 3 as a threshold, which indicates significant co-occurrence of collocations (e.g. Hunston, 2002; Vilkaite, 2016; Wolter & Yamashita, 2015). Since the MI-score is an effect size measure, it does not take how much evidence the observed corpus provides into consideration. Therefore, it is questionable that MI-score of 3 indicates significance co-occurrence of collocations.

A number of corpus-based studies explored the co-selection of words by L1 English and L2 learners of English (e.g. Durrant & Schmitt, 2009; Bestgen & Granger, 2014) employing the largely used AMs, t-score and MI-score. In order to see the contributions of these AMs to the corpus-based SLA research, it is useful to briefly review these studies. Durrant and Schmitt (2009) investigated the extent to which L1 and adult L2 speakers of English use collocations in their writings. They extracted directly adjacent (n-gram) 10839 adjective-noun word combinations from L1 and L2 speakers' writing. As AMs, they employed t-score and MI-score. They took t-score of 2, and MI score of 3 as minimum conditions of collocations. They found that low-frequency, MI-score based collocations were more prevalent in L1 than L2 speakers' writings. That is to say, L2 speakers underused MI-based collocations. They also found a significant overuse of high-frequency, t-score based collocations in L2 speakers' writings. Another corpus-based study conducted by Bestgen & Granger (2014), using a longitudinal approach, explored whether L2 learners' use of collocations in writing develop over time. They used the Michigan State University corpus of English as a second language writing made up of 171 essays written by adult learners of English, and the Corpus of Contemporary American English (COCA) as a reference corpus. They used t-score and MI-score as AMs to investigate L2 learners' use of collocations. They found that the number of high-frequency t-score based collocations decreased gradually. However, no significant increase was observed in the number of low-frequency MI-score based collocations in L2 learners' writings.

To highlight the issues in the use of AMs by corpus-based SLA studies, it is important to revisit one of their main findings (e.g. Durrant & Schmitt, 2009). They repeatedly found that L2 users produced fewer MI-based collocations than L1 users (Durrant & Schmitt, 2009, Bestgen & Granger, 2004, Schmitt, 2012). In other words, these studies reported that L2 users produced more frequent collocations, as they appear to occur frequently in a reference corpus (e.g. BNC), but these collocations were not strongly associated according to their MI-scores. As previously discussed by a number of studies (e.g. Evert, 2008; Durrant and Schmitt 2009; Gablasova et al. 2017; Schmitt, 2012), the MI-score does not only highlight the exclusivity in the collocations, but also it is negatively linked to high-frequency of the word pairs. Therefore, it is important to pay attention to low-frequency aspect of MI-score, as well as highlighting its exclusivity. As Schmitt (2012) also pointed out that MI-score highlights infrequent collocations whose constituents may also be low-frequency words themselves, and the collocations may be "more specialised" involving technical terms (Ebeling & Hasselgård 2015, p. 211). At this point, Gablasova et al. (2017) suggests that these less frequent, and more specialised collocations (e.g. *densely populated*), highlighted by MI-score, may not be known by the L2 learners yet. Therefore, using only MI-score based collocations might not be the most ideal way for measuring L2 speakers' collocational knowledge or processing. We therefore need alternative ways of operationalising collocations to explore L2 collocational knowledge.

Kang's (2018) study is one of the rare studies which compared a few AMs to explore the relationship between word association and collocations in a general corpus. He compared the word associations in the Edinburgh Association Thesaurus (EAT), and collocations in the BNC. The participants of the EAT experiments were asked to provide the first word that came to their minds after seeing the stimulus words. The participants were 100 presumably L1 speakers of English undergraduates. After excluding very low frequency stimulus words, Kang (2018) used 3177 stimulus words as the nodes to search for collocations. To measure collocational strength, he used t-score, MI, local-MI, and simple log-likelihood. For each stimulus word in the EAT, he firstly checked the primary response word for its rank among the collocates in the BNC using the above-mentioned AMs. These ranks of primary response were later used to determine the strength of the relation between word association and collocation for each of the AMs used. He found that the AM which mirrored word association the closest is the simple ll. The average rank of simple ll was larger than that of other methods. The MI-score showed the poorest performance in replicating the word association. Furthermore, t-score performed better than the MI-score. Based on this finding, Kang (2018) concluded that significance testing measures (e.g. t-score and simple ll) perform better than the effect size measures in reflecting the relation between word association and collocations. This conclusion with regard to significance testing measures' predicting the word association better than effect size measure could be slightly premature because these findings could also possibly be affected by the choice of particular effect size measure. As previously pointed out, the MI-score highlights infrequent word pairs with more specialised meanings (Ebeling & Hasselgård, 2015, Gablasova et al. 2017). However, it is likely that collocational frequency might have affected the participants' responses for word association task. In this regard, the use of an effect size AM which is not negatively linked to high-frequency could have predicted the word association better.

Considering the shortcomings of MI-score and t-score, Gablasova et al. (2017) introduced the Log Dice score (LD-score) as an alternative measure of association. The LD-score is calculated as the harmonic mean of two proportions, which indicates the tendency of two words to co-occur relative to the frequency of these words in the corpus (Evert, 2008, Smadja, McKeown, & Hatzivassiloglou, 1996). The LD-score is a standardised measure based on a scale with a fixed maximum value of 14, it is therefore a comparable measure across different corpora. In addition, with LD-score, it is possible to see that how far the value of particular word pair is away from the theoretical maximum, which indicates an entirely exclusive collocation (Gablasova et al. 2017). A negative LD-score show that the co-occurrence frequency is lower than what would be expected by chance. Unlike the MI-score, the LD-score is not negatively linked to low-frequency. In this regard, LD-score is preferable to the MI-score especially if the goal of the researchers is to explore the exclusivity between the words in the collocation without punishing for the high-frequency (see Gablasova et al. 2017, Rhycly, 2008, for a more detailed discussion on LD-score). Since LD-score does not punish for high-frequency, it can be used to extract both high-mid-frequency and exlusive collocations that can be used to explore L2 collocational knowledge and processing. Gablasova et al. (2017) explored how much variation the three measures (t-score MI-score, and LD-score) cause across sub-corpora of the BNC by comparing the association scores of three verb-noun collocations (*make sure, make decision, make point*). They found that the t-score varies across corpora considerably more than MI and LD scores because of its strong reliance on corpus size. Therefore, it is difficult to compare the association scores of collocations across different corpora based on t-score.

It should be noted that the directionality is a property of collocations. However, none of the AMs discussed above (t-score MI-score, and LD-score) view collocations as entirely symmetrical combinations, and they do not distinguish whether the first word is more predictive of the second word or the other way around. Delta P measure, arising out of associative learning theory, captures the directionality of associations in word pairs (Gries, 2013). Delta P produces two different values of association for any pair of words and thus it is possible to observe which word predicts the other in a two-word collocation. For example, in adjective-noun collocations it is possible to identify whether the adjectives or the nouns are more predictive of the following or preceding words within collocations. Delta P shows the probability of the outcome given the cue versus the probability of the outcome when the cue is absent. It is a standardised measure based on a scale with a fixed minimum and maximum values. Delta P approaches 1.0 as the presence of the cue increases, the probability of the outcome and approaches -1.0 as the cue decreases the probability of the outcome (see Gries, 2013, p. 143-152 for a more detailed discussion on Delta P). Since Delta P is not a significance test, it is not affected by corpus size, which makes it a comparable measure across corpora.

3.2 Contrastive studies of MWS in different languages

Before reviewing the corpus-based contrastive studies, it is important to clarify the term "contrastive" in corpus-based research of MWS. Colson (2008) suggests that it is possible to interpret the term "contrastive" in more than one way. One possible interpretation of the term "contrastive" is to see it as a synonym of "crosslinguistic". In this regard, any type of comparison of set of phrases such as a comparison of idioms or lexical bundles with their translated versions in another language is considered as "contrastive phraseology" (e.g. Dayrell, 2007). However, it is also possible to interpret "contrastive" in a narrower sense, which implies a more systematic, in-depth corpus-based comparison of the use of a specific type of MWS in two or more languages. This section chose the latter approach. In other words, it reviews the contrastive corpus studies which aim to provide a systematic analysis of how a specific type of MWS (e.g. lexical bundles, or collocations) used in different languages.

Cortes (2008), using comparable corpora, aimed to identify and analyse the structures, and functions of lexical bundles in academic history texts. Two corpora were collected to identify the lexical bundles in published history writing in English and Spanish. Each corpus contained around one million words. Lexical bundles were operationalised by their frequencies. She extracted only four-word lexical bundles with frequency cut-off points of occurring twenty times in one million words and at least in five texts. For the functional analysis, she used Biber's et al. (2004) taxonomy. Cortes (2008) found that the number of lexical bundles in the corpus of Spanish history writing (183) was more than twice as large as found in the English corpus (87). Since Spanish nouns, pronouns, and demonstratives are marked for gender, and number, those lexical bundles in the Spanish corpus showed this variation but belonged to the same root were grouped together. The result of this process did not make a considerably difference, a total of (163) Spanish bundles remained. Cortes (2008) argued that nouns are predominantly modified by adjectives and pre-positional phrases in Spanish. While, noun-noun pre-modification is very frequent in English academic writing (e.g. *immigration history*), it is not possible in Spanish. It needs to be expressed by a post modifying prepositional phrase (e.g. *la historia de la immigración*). Therefore, there are more four-word lexical bundles in Spanish corpus than English corpus. Despite the large differences in the number of lexical bundles in Spanish and English corpora, there is a high level of agreement among the structural types. The most frequent category of lexical bundles in both corpora were prepositional phrases (e.g. *as a result of*). The second most frequent category in both corpora was noun phrases (e.g. *the beginning of the*). Furthermore, lexical bundles identified in both corpora fulfilled similar functions. The most frequent category was referential bundles in both English and Spanish.

Granger (2014) investigated the use of lexical bundles in English and French. Since quality and quantity of the lexical bundles are inevitably affected by genre. She chose to conduct the study based on a matched corpora representing two genre – parliamentary debates (spoken) and newspaper editorials (written). The matched subcorpora were extracted from the Europarl, a multilingual corpus of proceedings of the European parliament debates, and the Mult-ed, a multilingual corpus of editorials from quality papers. She extracted two one-million-word of original texts in English and French from both corpora. She chose 3-to 7-word bundles from various frequency bands using the cut-off points (50, 40, 30, 20, 10, 5, and 3) for each language and genre. Granger (2014) found that there is overall a higher number of lexical bundles in French editorial texts than that of English. She reported that for the parliamentary debate genre the picture looks more complex. The frequencies of the bundles were slightly higher in French than in English. However, English displayed a higher number of stem bundles than French (e.g. *it is clear that*). Particularly, English corpus of parliament debates appeared to contain a higher number of stem bundles with 1st person singular (e.g. *I am delighted that*) than French corpus of parliamentary debates. Granger (2014) concluded that there are marked differences between the set of bundles found in two genres in both English and French. Furthermore, both languages appear to be comparably formulaic. One shortcoming of this study is that it is not clear on what basis, Granger (2014) established the cut-off points (50, 40, 30, 20, 10, 5, and 3) for frequency bands to extract the lexical bundles.

In addition to lexical bundles, there has also been only little research on corpus-based contrastive studies of collocations. To address this gap, Xiao and McEnery (2006) undertook a contrastive analysis of collocation, semantic prosody and near synonymy in English and standard Chinese. They used the Freiburg-LOB corpus of British English, Freiburg-Brown corpus of American English, and the Lancaster Corpus of Mandarin Chinese. Each of these corpora contained approximately one million words collected from fifteen written text categories. As supplementary corpora, they also used the LOB corpus of British English, the Brown University Corpus of American English, and the People's Daily Corpus for Chinese. They specifically explored whether Chinese exhibits semantic prosody, and semantic preference as English does, and how different are the collocational behaviour, and semantic prosody of lexical items with similar denotational meanings. They employed MI-score of 3 as a cut-off point to extract collocations, set the minimum co-occurrence frequency to 3 within 4:4 window span. They used Stubb's (2002: 225) definition of semantic prosody which is the meaning arising from the interaction between a given node and its typical collocates. The primary function of the semantic prosody is to express speaker/writer's attitude (Louw, 2000). Xiao and McEnery (2006) specifically focussed on the words *consequence*, *cause*, *price/cost*. They found that while English and Chinese are very unrelated languages, there are many similarities between collocational behaviours and semantic prosodies of near synonyms in the two languages. For example, the verb *cause* has overwhelmingly negative semantic prosody in both English and Chinese. However, some differences were also observed in the two languages. For example, the four near synonyms of the noun *consequence* in English can be placed on a semantic continuum from positive to negative as follows: *outcome/result, consequence and aftermath.* In Chinese, semantic prosodies of the near synonyms of *consequence* appeared to be more sharply divided between clearly negative, and clearly positive at the ends of the continuum. Xiao and McEnery (2006) concluded that a more general difference between the two languages that collocation and semantic prosody might be affected by the morphological variation in English, but not in Chinese, since it lacks such variation.

The contrastive corpus-based studies reviewed in this section played an important role in providing corpus-based evidence for revealing the similarities and differences in the use of collocations and lexical bundles in different languages. It should be noted that there is only little contrastive corpus research on collocations and other type of MWS focussing how and in what frequency they are used in different languages. In order to be confident that MWS are pervasive and universal feature of language, we need more evidence from contrastive corpus studies on MWS. Furthermore, studying MWS using a contrastive approach would increase our knowledge about MWS, and in what ways they are affected by language typology.

3.3 MWS in agglutinating languages

As emphasised in section 2.3, the study of MWS has focussed only on a narrow range of languages, especially on English, and this approach has inevitably restricted the scope of formulaic approach to a few selected languages (Durrant, 2013). Biber (2009) argued that agglutinating languages are interesting field of exploration (e.g. Turkish and Finnish), which rely on extensive system of suffixes to create complex word forms. Exploring formulaicity in agglutinating languages is different from that in non-agglutinating languages since their rich morphology opens up the possibility that formulaicity might take place not only between the words but also within the words. Unsurprisingly there has been very little corpus-based research focussing on the formulaicity within and between words in agglutinating languages.

To address this research gap, Durrant (2013) conducted a pioneering corpus study to explore the within-words formulaic patterns in Turkish. He aimed to provide a description of syntagmatic associations between items, fixed sequences of items, and associations between particular lexical and grammatical forms at the morphological level. He collected a corpus of Turkish newspaper articles consisting of 515 news items, and 250 opinion pieces – totalling 374,590 words. He looked at the formulaic patterns within verbal inflections, and analysed 20 verbs in total. He selected the verbs from different frequency bands. He referred the co-occurrences of suffixes as "collocational relationships between suffixes". Durrant (2013), focussing on collocational relationships between suffixes, found that most high-frequency suffixes enter into both novel and regular combinations, and the strongest collocations of suffixes consist of two immediately adjacent suffixes. Furthermore, he found a considerable variation in the types and strengths of collocational relationship into which they enter. In order to analyse longer suffix sequences, he generated a list of the most frequent, three- and four-morpheme bundles. Durrant (2013) found that some very high-frequency three-morpheme bundles, which are used across a wide range of verb roots. They were dominated by two types of structures. Eighteen out of twenty combinations included subordinators plus person markers, while the remaining two items had negative forms. Finally, Durrant (2013) looked at the spread of three-morpheme suffix combinations across verb roots. He found that while the three-morpheme bundles are available for use across most verb roots, the suffix combinations appeared to have stronger associations with some verb roots. Durrant (2013) concluded that high-frequency morphological patterns can be found within Turkish words, which provided evidence that formulaic patterning is not limited to word-level. Furthermore, Durrant (2013) suggested that any model of language which views grammar and lexis as independent systems is inadequate to describe the formulaicity of agglutinating languages. However, a usage-based model of morphology which takes frequency of the suffix combinations into consideration is more consistent with the formulaic patterns within Turkish words. To the best of my knowledge Durrant's (2013) is the only corpus-based study focussing on the formulaicity of Turkish.

Similar to Durrant (2013), this thesis also explores the formulaicity of an agglutinating language. However, this thesis differs from the previous works in some respects. First, the present study primarily focuses on the formulaicity between words rather than the within words. Second, this study contrastively investigates frequency and association statistics of collocations in English and Turkish. This enables us to see the extent to which language typology affects the frequency and association of collocations (see Chapter 4 for corpus study). Having seen the findings of the corpus study, this thesis also presents psycholinguistic experiments aiming to ascertain the extent to which L1 English and Turkish participants rely on the same mechanisms for processing adjective-noun collocations. Furthermore, the present thsis explores whether the participants' response times (RTs) for adjective-noun collocations mirror the patterns emerged from the corpus study, in relation to the collocations' frequency of occurrence and collocational strength. Moreover, the present study investigates whether the same factors affect L1 English and advanced L1 Turkish-English L2 learners' processing of collocations (see Chapter 6 for the psycholinguistic experiments)

Chapter 4: A Contrastive Corpus Study of Adjective-noun Collocations in English and Turkish

This study examined frequency counts and association statistics of adjective-noun collocations in English and Turkish from a contrastive perspective. Although the importance of frequency and association counts of collocations are well-recognised by corpus-based SLA, and psycholinguistic studies (Durrant & Schmitt, 2009; Bestgen & Granger, 2014; Kang, 2018; Wolter & Gyllstad, 2011; Wolter & Yamashita, 2015), there has been only little work done on the different properties of collocations using AMs in languages other than English (e.g. Xiao & McEnery, 2006). In addition, to the best of my knowledge, no previous study has contrastively explored the formulaicity of agglutinating and non-agglutinating (analytical) languages. To be able to extend our overall knowledge of MWS in general, and to develop fine-grained models of formulaicity of agglutinating languages, it is important to conduct contrastive corpus studies focussing on agglutinating and non-agglutinating languages. Therefore, this study, working with English and Turkish, aims to explore the ways in which language typology, particularly agglutinating structure of Turkish affects the collocability of words. Turkish, a language with rich agglutinating morphology, sometimes constructs complex word forms using extensive system of suffixes whereas English, as Biber (2009) noted, it has minimal inflectional morphology, but a large set of grammatical function words. Those two factors play an important role in the formulaic pattern types that are common in English. However, as Durrant (2013) noted, high-frequency morphological patterns can be found within Turkish words, which provided evidence that formulaic patterning is not limited to word-level in agglutinating languages. For instance, rich nominal morphology of Turkish, consisting of case, plural, instrumental and person marking suffixes, ensures that a Turkish noun can have more than ten different inflected forms (see section 2.3.1 for examples). It is therefore interesting to explore Turkish and English collocations contrastively to find out if agglutinating structure of Turkish has any impact on the collocability of words.

In the present study, frequency counts and association statistics of Turkish and English adjective-noun collocations are contrastively explored using various AMs with the aim of addressing two research questions (see also section 2.3.3 for research questions):

(i) How different (or similar) are frequency counts and association statistics for translation-equivalent English and Turkish adjective-noun collocations?

Word combinations in Turkish are predicted to be less frequent than their equivalents in English. Since meanings which require multiple word expressions in English can be expressed using a single word in Turkish, individual word forms (lexemes) are expected to have lower frequency counts than their English equivalents.

(ii) How different (or similar) are the frequency and association scales for English and Turkish adjective-noun collocations?

Due to the agglutinating structure of Turkish, I would predict to find a notably larger difference between the unlemmatised and lemmatised Turkish collocations' frequency and association counts than English collocations' frequency and association counts.

The aim of the first research question is to investigate whether the agglutinating structure of Turkish affects the collocability of adjectives and nouns – also known as syntagmatic associations. In order to thoroughly investigate this, this study also compared lemmatised adjective-noun collocations to see if lemmatisation affects the frequency and association counts for English and Turkish collocations. Lemmatisation includes the frequency counts of both the base and inflected forms of the collocations. The aim of the second research question is to reveal the frequency and association scales of Turkish and English adjective-noun collocations because these scales will provide important references for how to select collocations for follow-up psycholinguistic experiments looking at processing of English and Turkish collocations. To be able to work with English and Turkish collocations, the type of collocations needed to be carefully chosen to study the two languages contrastively. This study chose adjective-noun collocations to contrastively explore English and Turkish MWS. The first reason for choosing adjective-noun collocations is that nouns within adjective-noun collocations can be inflected with various types of suffixes including case marking, plural and instrumental in Turkish, and thus it is possible to observe the influence of agglutination on the collocability of adjectives and nouns. Second, they occur in a certain syntactic order in which adjectives precede the nouns in both Turkish and English, hence they should be fully comparable for both strength and directions of the associations in Turkish and English. It sholed be noted that verb-noun collocations might have shown even stronger effects with regard to agglutinating structure of Turkish because Turkish verbs are very rich in terms of morphology (see Durrant, 2013). Therefore, they are more likely to affect the collocability of verbs and nouns. Nevertheless, important advantages of using adjective-noun collocations are that they follow the same syntactic order in the both languages and the uninflected (base) forms of the collocations can be directly compared in English and Turkish.

4.1 Method

4.1.1 Corpora.

Corpora chosen for this chapter need to represent the input that Turkish and English language users experience on a daily basis. The Turkish National Corpus (TNC) a written and spoken general corpus of Turkish with a size of 50,678,199 tokens (approximately 2 million tokens of spoken component), excluding punctuation marks, was used to investigate the adjective-noun collocations' frequency of occurrence and association counts in Turkish (Aksan, Aksan, Koltuksuz, Sezer, Mersinli, Demirhan, Yilmazer, Atasoy, Öz, Yildiz, Kurtoglu 2012). The TNC is a collection of 4,438 different text samples, representing 9 domains and 39 different genres, written in between 1990-2013. Given the nature of this corpus, the TNC can be considered a balanced and large corpus of modern written Turkish. The British National Corpus (BNC) XML edition was used to investigate the adjective-noun collocations' frequency of occurrence and collocational strength in English. It is a written and spoken corpus with a size of 98,560,118 tokens in 4,048 different text samples, excluding punctuation marks (approximately 10 million token of spoken component) (Burnard, 2007). Although the BNC is slightly dated since it includes the texts between 1960-1993, it is still largely considered a large and balanced corpus of general English. This study included both written and spoken components of the both corpora. Ideally, I would prefer to use a Turkish corpus of a bigger size for comparability purposes. However, the fact that TNC is a fairly representative corpus of modern Turkish and the distributions of genres are very similar in the two corpora.

It is questionable that to what extent those large-scale corpora such as the TNC and BNC can be representative of the input that language users experience on a daily basis. Although corpora are usually designed to include balanced sample texts of a particular language domain such as a national variety (e.g. BNC), language users are less likely to be exposed to the full range of different styles present in a large-scale general corpus (Durrant, 2013). In line with this, Hoey (2005) acknowledged the mismatch between the ranges of texts corpora contain and language users' daily experiences of the language. However, he also emphasised that corpora have a certain potential to present a type of language that language users might encounter. Some other researchers argue that there is still a reason to be suspicious of research drawing conclusions about individuals' experience particularly L2 speakers on a large general corpus (Durrant, 2013, González Fernández & Schmitt 2015). Having acknowledged the potential mismatch between the range of different types of language (registers) large scale general corpora contain, and language users' daily experiences of the language, this study uses the BNC and TNC, as a useful proxy of the target language; in this way, it is possible to look at frequent and widely distributed collocations in both languages. It should be noted that the follow up psycholinguistic experiments (see Chapter 6) will provide complementary empirical evidence about the extent to which L1 and L2 speakers are sensitive to the adjective-noun collocations extracted from large scale general corpora the BNC and TNC.

4.1.2 Procedure.

In this study, corpus-based analysis followed the node-collocate approach to collocations. Nodes are the words of interests and the collocates are the words occurring around the node words (Ebeling & Hasselgård, 2015). In order to extract the adjective-noun collocations, firstly, the frequency bands were established making use of the TNC and the BNC word frequency lists. Frequency distributions of the two corpora largely follows the Zipfian distribution, with a very small number of high-frequency nouns and a long tail of low-frequency nouns. In the BNC the most frequent noun *time* occurred with the relative frequency of 1,842.00 per million words (PMW) as a lemma, in the TNC the most frequent noun *iç* (the inside) occurred with the relative frequency of 3,362.58 PMW as a lemma. The infrequent nouns such as *assister*, *leaser*, *lox, dentin* occurred with the relative frequency of 0.010 PMW as lemmas in the BNC. The infrequent nouns such as *firek* occurred with the relative frequency of 0.02 PMW as lemmas in the TNC. With that in mind, for high-frequency words 400 or above, for mid- frequency words between 150 and 350, and for low frequency words 100 or below relative frequency were determined as cut-off values to establish the frequency bands. Following the node-collocate approach, a total of forty adjective-noun collocations were extracted from high-mid-frequency bands. Low-frequency band collocations were not within the scope of this study, which intended to include collocations that language users frequently encounter. In the corpus study the collocation frequency was treated as a categorical variable to be able to compare the collocations in different frequency bands in English and Turkish. for comparability purpose, only the nouns within the same frequency bands in the BNC and the TNC were chosen as nodes. Using the CQPWeb tool (Hardie, 2012), the most frequent four adjective collocates of each node words were extracted in the BNC and their equivalents in Turkish were identified through the TNC interface. This study used L3-R3 window span.

Table 4.1 Node words

Frequency band	Node words (Turkish)	Node words (English)	Relative node frequency in the TNC	Relative node frequency in the BNC
High-frequency	Zaman	Time	1690	1360.38
400+	Gün	Day	1091.06	535.57
	Dünya	World	605.01	512.45
	Yol	Way	512.51	853.69
	Adam	Man	490.8	524.06
Mid-frequency	Ülke	Country	263.31	279.72
150-350	Aile	Family	235.8	300.56
	Toplum	Society	231.68	209.54
	Sonuç	Result	195.63	195.45
	Sanayi	Industry	170.94	176.12

This corpus-based contrastive study selected only the most frequent four adjective collocates of each node words from the BNC. Then the translation-equivalents of the collocations were extracted from the TNC. If more than one adjective collocate in Turkish can possibly be the equivalent of an English collocate, the more frequent collocate was selected as the equivalent. If the meaning-equivalent of a possible adjective collocate in English is not classified as an adjective in Turkish, the collocate was discarded from the analyses. For the frequency comparisons of adjective-noun collocations in English with their equivalents in Turkish, raw frequency scores in the BNC and the TNC were relativized to PMW. This allows a comparison of how many times a Turkish collocation is likely to occur PMW against its equivalent in English. The BNC XML edition was tagged by CLAWS (Rayson & Garside, 1998). The collocational lists of each inflected forms of the node words in Turkish and English were searched for whether they collocate with the adjectives in search through the TNC interface and CQPWEB tool for the BNC. If they collocate, the raw and relative frequency scores of the nodes and the collocations were extracted from the TNC and the BNC. Finally, those values were listed in a spreadsheet as shown in sample demonstration in Tables 4.2 and 4.3.

Table 4.2 Base and Inflected forms in English

Node + collocate Good time	Raw frequency of the node in the BNC	Relative frequency of the node in the BNC	Raw collocate frequency in the BNC	Relative collocate frequency in the BNC
Good time (base form)	152502	1360.38	1138	10.15
Good time-s (pl)	29194	260.42	251	2.23

Note. Both single word and collocation frequency scores were extracted from the BNC.

In Turkish the base, five case marking (accusative, dative, genitive, locative, ablative), and plural inflected forms of the node words were investigated for the collocations' frequency of occurrence and associations between the nodes and the collocates as shown in Table 4.3.

Table 4.3. Base and Inflected forms in Turkish

Node + collocate Iyi zaman (Good time)	Raw frequency of the node	Relative frequency of the node	Raw collocation frequency	Relative collocation frequency
Iyi zaman (base form)	85646	1690	765	111.32
Iyi zaman-I (Acc)	5163	101.88	94	6.71
Iyi zaman-A (Dat)	2953	58.27	24	3.83
Iyi zaman-In (Gen)	3326	65.63	25	4.32
Iyi zaman-Da (Loc)	16243	320.51	125	21.11
Iyi zaman-DAn (Abl)	1207	23.82	0	
Iyi zaman-lAr (Pl)	6019	118.77	45	7.82

Note. Both single word and collocation frequency scores were extracted from the TNC.

It should be noted that the nodes with more than one suffix such as case marking and plural at the same time were discarded from the analysis because they are very low-frequency word pairs. Unfortunately, to the best of my knowledge it is not possible to find out how much data was lost in this way because the TNC does not support this type of searches. If the difference in relative collocate frequency (hereafter, RCF) scores was considerably large between English and Turkish, the concordance lines were consulted to understand the nature of the difference.

Besides the frequency comparisons, this study looked at the association scores between the nodes and the collocates, also known as the collocational strength as measured by MI, LD, and Delta P measures. The collocational relationship is a complex one and no single measure of association can capture the full complexity of this relationship (Brezina et al. 2015). Therefore, this research employed three different corpus derived AMs to investigate the collocational strength between the nodes and the collocates of two-word adjective-noun combinations in English and Turkish. As discussed previously (see section 3.1), the MI-score tends to highlight relatively infrequent words with low co-occurrence frequency (Evert, 2008, Manning & Schütze, 1999). Thus, another measure of association, which is neutral to the low-frequency of occurrence, was needed to explore association scores (Gablasova et al. 2017, Rychly, 2008). In this regard, LD measure was also used in this study. MI and LD measures consider collocational strength as symmetrical. Therefore, Delta P measure was also used as a measure of association alongside these measures to investigate the directions of the associations. To be able to compare the strength of associations between the nodes and the collocates in Turkish and English as measured by these scores, the following values were directly collected from the TNC and the BNC.

i. Number of tokens in the whole corpus: N
ii. Frequency of the node in the whole corpus: R_1
iii. Frequency of the collocate in the whole corpus: C_1
iv. Frequency of the collocation (i.e. node + collocate) in the collocation window: O_{11}

The wide range of complex inflections in Turkish means that many word forms may occur with very low frequencies and thus it becomes quite difficult to make generalisations at the lexical level (Durrant, 2013). Therefore, lemmatising the inflected forms to abstract away from the complex morphology seems to be a natural step. Lemmatisation is defined as grouping together word forms that belongs to the same inflectional morphological paradigm and assigning to each form its canonical form, which is called headword (Gesmundo & Samardžić, 2012). For calculating the lemmatised RCF and association scores in Turkish and English, the raw frequency counts of all of the inflected forms of the nodes and all of the inflected forms of the collocations were identified through the TNC interface and the CQPWEB tool for the BNC. The frequency sums of the base and inflected forms of the node words were taken as frequency of the node in the whole corpus (R_1) and the frequency sums of the base and all inflected forms of the collocations were taken as frequency of the collocation in the collocation window (O_{11}). Thus, RCF and association scores were calculated for both inflected and uninflected forms of collocations in Turkish and English. After calculating the RCF and association scores for each inflected form, the same measures were calculated for the lemmatised forms of the collocations. If RCF or association scores were considerably large between unlemmatised and lemmatised forms of the collocations, the concordance lines were consulted to understand the possible reasons of the large differences.

In addition to comparing frequency and association counts of adjective-noun collocations in English and Turkish, this study aimed to identify the adjective-noun combinations' scales of association counts as measured by LD scores. Firstly, adjective-noun combinations with lowest and highest frequency and association scores were detected in the high- and mid-frequency bands for RCFs and each AMs. In order to find out the highest and lowest association scores in each frequency bands, adjective collocates of each noun nodes in high- and mid-frequency bands were checked in English and Turkish through the lens of the MI and LD scores. Secondly, the percentages of non-collocates, which has negative association scores, weakly and strongly associated word combinations according to MI and LD measures were calculated in each frequency bands for English and Turkish. Therefore, it was possible to compare association scales of adjective-noun combinations in both languages. The scales were identified for L3-R3 window span, and the minimum collocation frequency was set to 5 for both languages and frequency bands. The scales were identified for non-lemmatised collocations only. The data analysis consisted of two main stages: comparing frequency and association scores of English and Turkish collocations and identifying the scales of adjective-noun collocations in English and Turkish (see Table 4.4 for a detailed overview data analysis).

Table 4.4 Overview of data analysis

1)	Comparing frequency and association scores of English and Turkish collocations
	Identifying the nodes and collocates
	Identifying the inflected forms of the collocations
	Calculating the unlemmatised RCF and association scores
	Calculating the lemmatised RCF and association scores
2)	Identifying adjective-noun collocations' scales of association and frequency
	Identifying the lowest and highest frequency and association scores in each band
	Calculating the percentages of weak and strongly associated collocations

4.2 Results

4.2.1 Comparing frequencies and associations of English and Turkish collocations.

In the high-frequency band, 20 adjective-noun pairs were extracted from the BNC using five node words; *time, world, man, way, day*, and equivalents of these collocations were identified in the TNC.

Four adjective-noun combinations extracted through the node *time*; *short time* (*kısa zaman*), *long time* (uzun zaman)*, good time* (iyi zaman)*,* and *right time* (doğru zaman) were investigated for frequency and association counts in the BNC and TNC. As shown in Figure 4.1, the RCF scores showed that two out of four unlemmatised combinations occur at higher frequency in English than their Turkish equivalents. The unlemmatised collocations *short time* and *long time* reach higher RCF scores of 13.82, and 46.82 in English than their Turkish equivalents *kısa zaman* and *uzun zaman* (11.4 and 27.03 respectively). The unlemmatised word combinations *iyi zaman*, and *doğru zaman* obtain higher RCF scores of 15.09, and 8.97 in Turkish than their English equivalents *good time* and *right time* (11.24 and 6.69 respectively). When these word combinations are lemmatised, all four combinations *kısa zaman*, *uzun zaman*, *iyi zaman*, and *doğru zaman* reach higher RCF scores of 49.25, 54.63, 26.24, 19.55 in Turkish than their English equivalents *short time, long time, good time* and *right time* (14.09, 47.19, 13.75, 6.88 respectively).

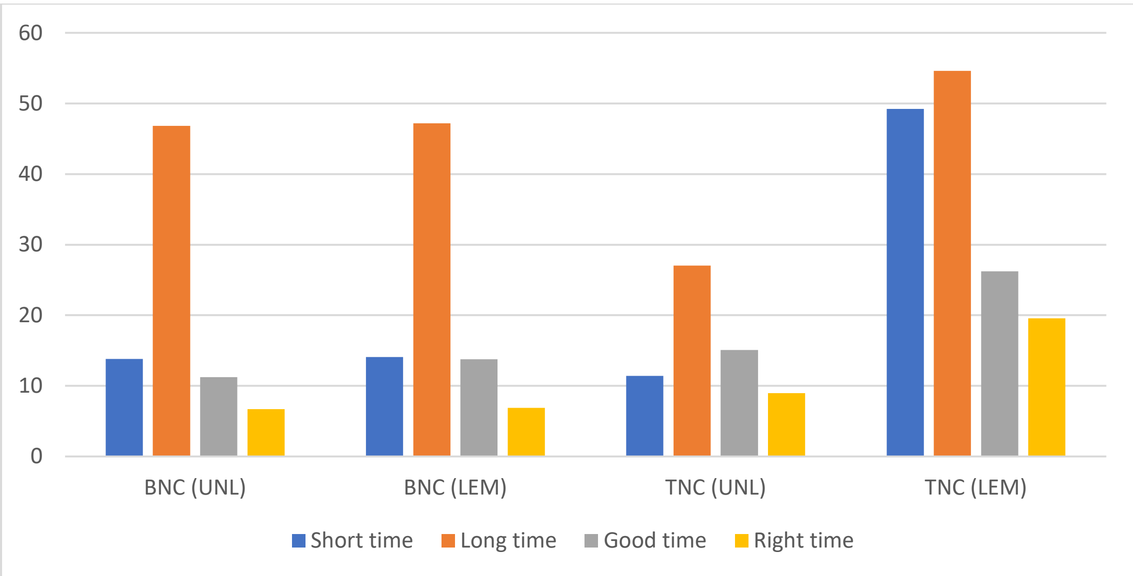

Figure 4.1 Relative collocation frequency scores (node time)

As shown in Figure 4.2, the MI-scores revealed that three out of four unlemmatised combinations reach higher MI scores in English than their Turkish equivalents. The combinations *short time*, *long time* and *good time* obtain higher MI-scores of 2.9, 3.15, 0.56 in

English than their Turkish equivalents *kısa zaman*, *uzun zaman*, and *iyi zaman* (1.41, 1.91, 0.19 respectively). The adjective-noun combination *doğru zaman* obtain slightly higher MI-score of -0.053 in Turkish than its English equivalent *right time* (-0.34). When these combinations are lemmatised, two combinations reach higher MI-scores in Turkish than their equivalents in English. The lemmatised combinations *kısa zaman*, and *doğru zaman* obtain higher MI-scores of 2.8, and 0.076 in Turkish than their English equivalents *short time*, and *right time* (2.67, and -0.55 respectively). The lemmatised combinations *long time* and *good time* reach higher MI scores of 2.91, and 0.6 in English than their Turkish equivalents (2.21, and 0.28 respectively). The LD-scores showed that the unlemmatised combinations *short time*, and *long time* obtain higher LD scores of 5.57, and 7.28 in English than their Turkish equivalents *kısa zaman* and *uzun zaman* (5.14, and 6.35 respectively). The unlemmatised combinations *iyi zaman* and *doğru zaman* reach higher LD scores of 5.43, and 4.73 in Turkish than their English equivalents *good time* and *right time* (5.18 and 4.42 respectively). When they are lemmatised *Kısa zaman*, *iyi zaman,* and *doğru zaman* reach higher LD scores of 6.56, 5.58, and 5.17 in Turkish than their English equivalents *Short time*, *Good time* and *Right time* (5.35, 5.24, 4.23 respectively).

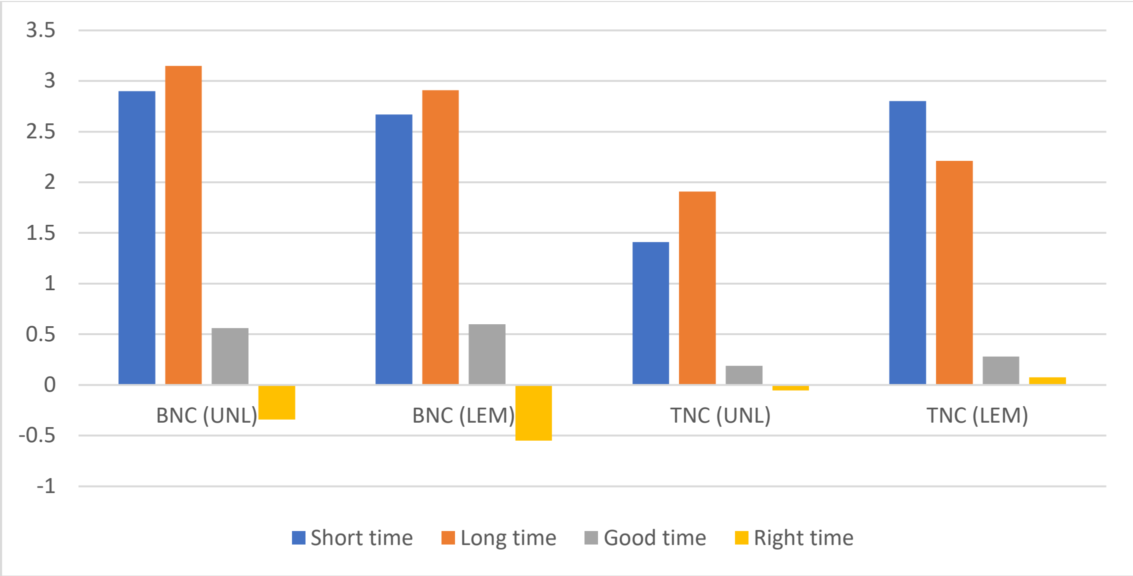

Figure 4.2 MI-scores (node-time)

The collocation *kısa zaman* (short time) is a notable example for the differences between lemmatised and unlemmatised RCF and association scores in Turkish and English. The unlemmatised collocation *short time* obtains RCF score of 13.82, and when it lemmatised its RCF slightly increase to 14.09 in English. Its Turkish equivalent, the unlemmatised collocation *kısa zaman* obtains RCF score of 11.4, when it is lemmatised its RCF score increases to 49.25. To understand the main reason for the sharp increase of the lemmatised collocation's RCF score in Turkish, I explored the inflected forms of the collocation. Its locative inflected form *kısa zamanda* occurs at a considerably higher frequency with RCF score of 36.11 than the base form of the collocation *kısa zaman* with RCF score of 11.4 in Turkish. Similar to its RCF scores, there are also large differences between the association counts of unlemmatised and lemmatised collocation *short time* in English and Turkish (as shown in Figures

Figure 4.2 and Figure 4.*3*). Therefore, I explored the association scores of 5 case inflected (accusative, dative, genitive, locative, and ablative), and plural inflected forms of collocation *kısa zaman* (short time) in Turkish. As in shown in Figure 4.3, shows that the difference between LD scores for unlemmatised and lemmatised collocation *short time* is not considerably different from each other in English (5.57 and 5.35 respectively). However, the difference between LD scores for unlemmatised and lemmatised collocation *kısa zaman* is considerably larger in Turkish (5.14 and 6.56 respectively). The reason for this radical increase should be the strongly-associated locative inflected form of the collocation *kısa zamanda,* which obtains LD-score of 8.97, because other inflected forms have considerably lower LD-scores than the base form of the collocation in Turkish.

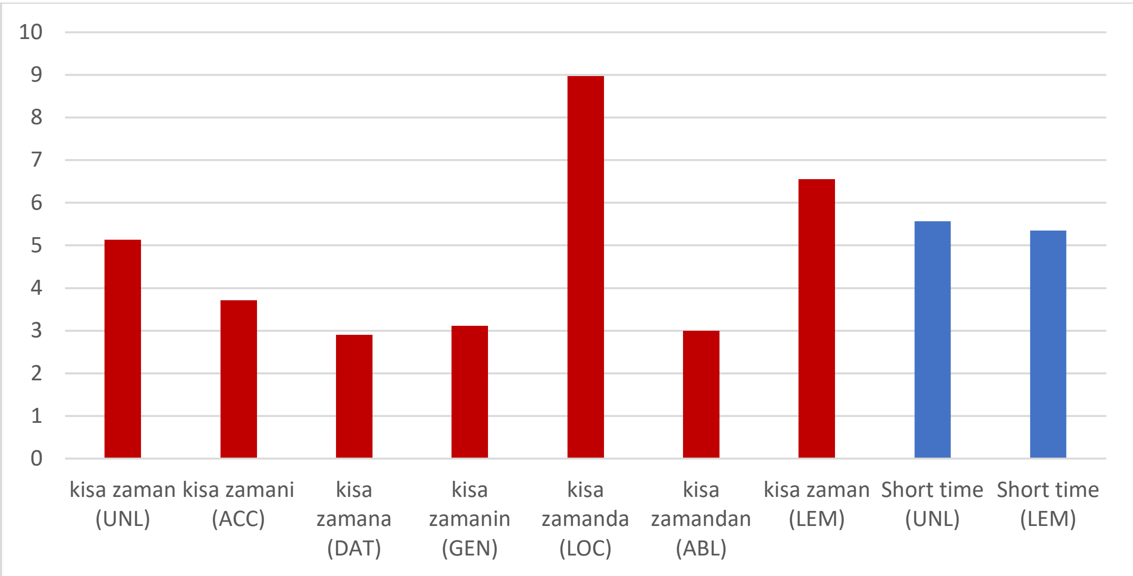

Figure 4.3 LD-scores for unlemmatised and lemmatised collocation short time

Four adjective-noun combinations extracted through the node *world*, *new world* (yeni dünya), *real world* (gerçek dünya), *outside world* (dış dünya), *whole world* (bütün dünya) were explored for frequency and association counts in the BNC and TNC. As shown in, the RCF scores showed that three out of four unlemmatised combinations occur at higher frequency in Turkish than their English equivalents. The unlemmatised combinations *yeni dünya, dış dünya,* and *bütün dünya* obtain higher RCF scores of 23.04, 7.3, and 16.25 in Turkish than their English equivalents *new world*, *outside world*, and *whole world* (8.32, 6.35, 5.08 respectively).

The combination *real world* reaches higher RCF score of 7.83 in English than its Turkish equivalent (2.2). When these word combinations are lemmatised, all four combinations *yeni dünya*, *gerçek dünya, dış dünya,* and *bütün dünya* reach higher RCF scores of 39.72, 9.47, 26.91, and 59.78 in Turkish than their English equivalents *new world*, *real world outside world*, *whole world* (8.5, 7.85, 6.38, 5.11 respectively).

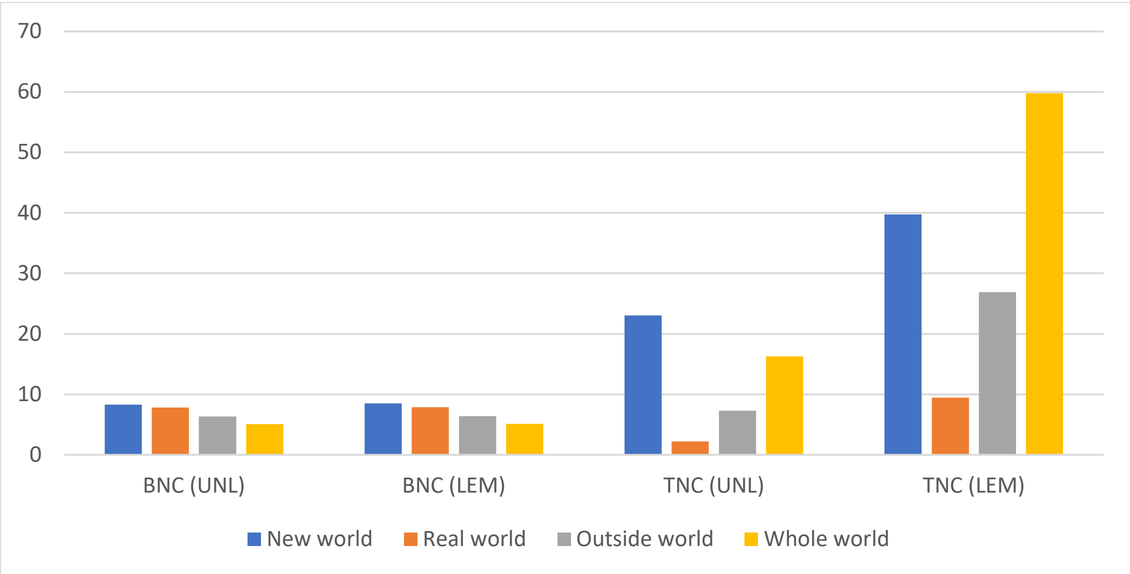

Figure 4.4 Relative collocation frequency scores (node-world)

As shown in Figure 4.5, the MI-scores showed that three out of four unlemmatised collocations reach higher MI scores in English than their Turkish equivalents. The combinations *real world*, *outside world* and *whole world* obtain higher MI-scores of 3.28, 3.08, 2.25 in English than their Turkish equivalent *gerçek dünya*, *dış dünya,* and *bütün dünya* (0.94, 2.57, and 1.88 respectively). The combination *yeni dünya* reaches higher MI-scores of 2.05 in Turkish than its English equivalent *new world* (0.92). When the combinations are lemmatised, the same combinations *real world*, *outside world* and *whole world* obtain higher MI-scores of 3.26, 3.06, and 2.42 in English than their Turkish equivalents *gerçek dünya*, *dış dünya,* and *bütün dünya* (0.94, 2.57, 1.88 respectively). The lemmatised combination *yeni dünya* reaches higher MI-scores of 1.41 than its English equivalent *new world* (0.92). The LD-scores showed that three out of four unlemmatised combinations *yeni dünya, dış dünya,* and *bütün dünya* have higher LD-scores of 7.19, 5.91, 6.78 in Turkish than their English equivalents *New world*, *Outside world* and *Whole world* (5.84, 5.81, and 5.45 respectively). The combination *real world* reaches a higher LD-score of 6.1 in English than its Turkish equivalent *gerçek dünya* (4.24). When they are lemmatised, three out of four lemmatised collocations *yeni dünya*, *dış dünya,* and *bütün dünya* reach higher LD-scores of 6.85, 6.45, and 7.48 in Turkish than their English equivalents *new world*, *outside world* and *whole world* (5.85, 5.79, and 5.43 respectively). The lemmatised collocation *real world* obtains a higher LD-score of 6.08 in English than its Turkish equivalent *gerçek dünya* (4.93).

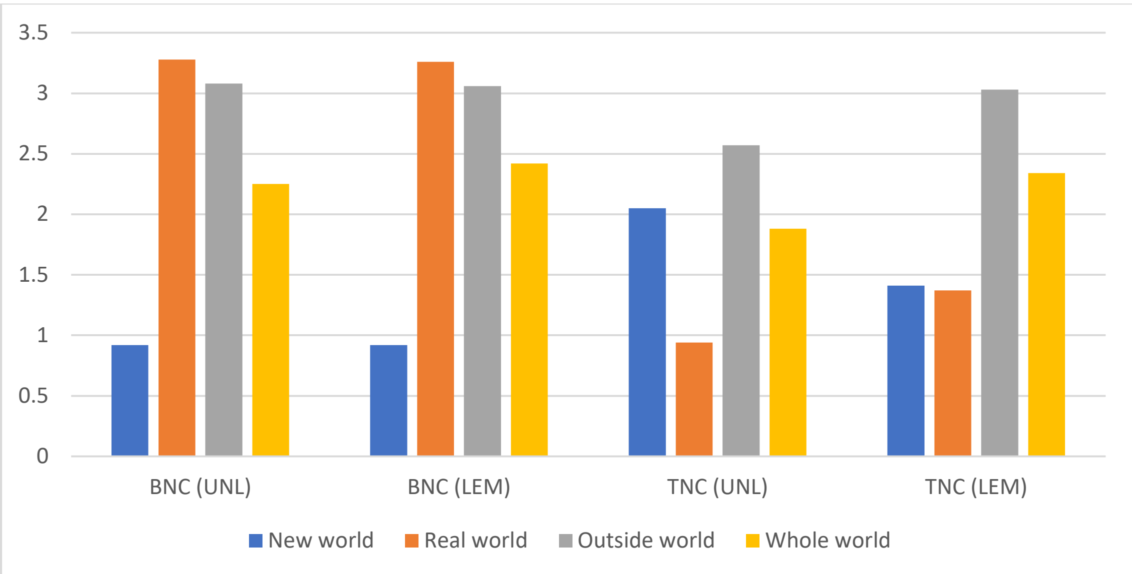

Figure 4.5 MI-scores (Node-world)

The collocation *new world* is another notable example for the differences between lemmatised and unlemmatised RCF and association scores in Turkish and English. The unlemmatised collocation *new world* has RCF score of 8.32 in English, and when it lemmatised, it slightly increased to 8.5. Its Turkish equivalent *yeni dünya* has RCF score of 23.4, and when it lemmatised, it increased to 39.72. The unlemmatised collocation *yeni dünya* reaches MI-score of 2.05, and LD-score of 7.19, when it is lemmatised both MI and LD scores decreased to 1.41, and 6.85 respectively. I explored the association counts of 5 case inflected (accusative, dative, genitive, locative, and ablative), and plural inflected forms of collocation *yeni dünya* (new world) in Turkish. As shown in Figure 4.6, the difference between LD scores for unlemmatised and lemmatised collocation *new world* is not considerably different from each other in English (5.84 and 5.85 respectively). However, the difference between LD-scores for unlemmatised and lemmatised collocation *yeni dünya* is considerably larger, and surprisingly the lemmatised collocations obtain lower association counts (7.19 and 6.85 respectively). The reason for this lower association counts for lemmatised collocation *yeni dünya* should be that the case and plural inflected forms do not appear to be strongly associated adjective-noun combinations, particularly the accusative, ablative, and plural inflected forms are considerably weaker associated than the base form Figure 4.6.

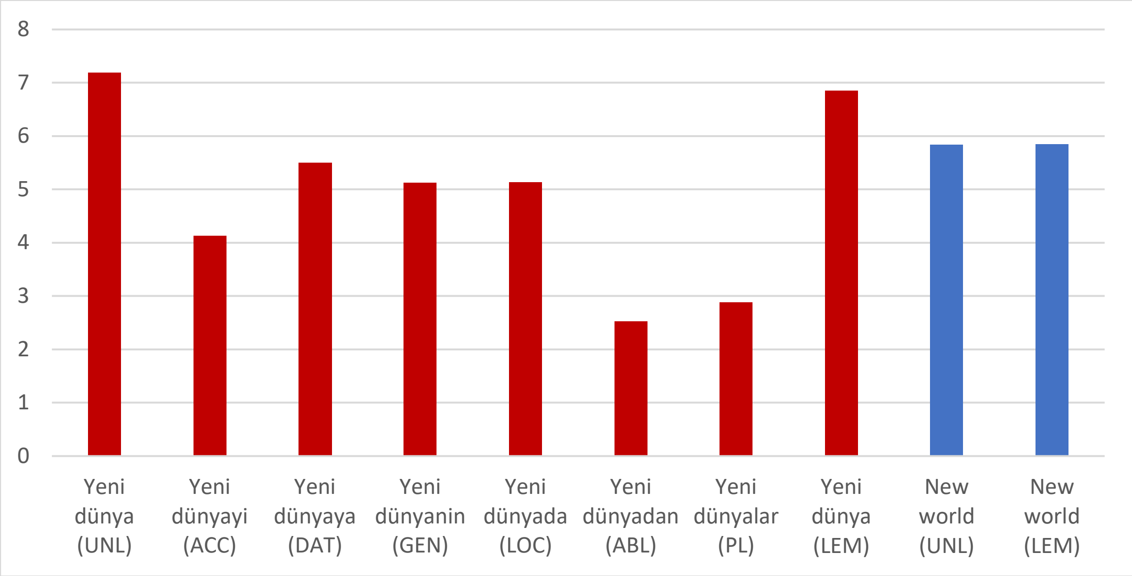

Figure 4.6 LD-scores for unlemmatised and lemmatised collocation new world

Four adjective-noun combinations extracted through the node *day*, *following day* (ertesi gün), *previous day* (önceki gün), *new day* (yeni gün), and *whole day* (bütün gün) were investigated for frequency and association counts in the BNC and TNC. As shown in Figure 4.7, the RCF scores showed that all four combinations occur at higher frequency in Turkish than their English equivalents. The unlemmatised combinations *ertesi gün*, *önceki gün*, *yeni gün*, and *bütün gün* reach higher RCF scores of 59.59, 28.43, 14, and 25.79 in Turkish than their English equivalents *following day*, *previous day*, *new day*, *whole day* (10.4, 4.34, 3.72, 2.65 respectively). When they are lemmatised, all four combinations *ertesi gün*, *önceki gün*, *yeni gün*, and *bütün gün* obtain higher RCF scores of 74.03, 34.66, 27.46, and 38.08 in Turkish than their English equivalents *following day*, *previous day*, *new day*, *whole day* (11.09, 4.97, 4.34, 3.08 respectively).

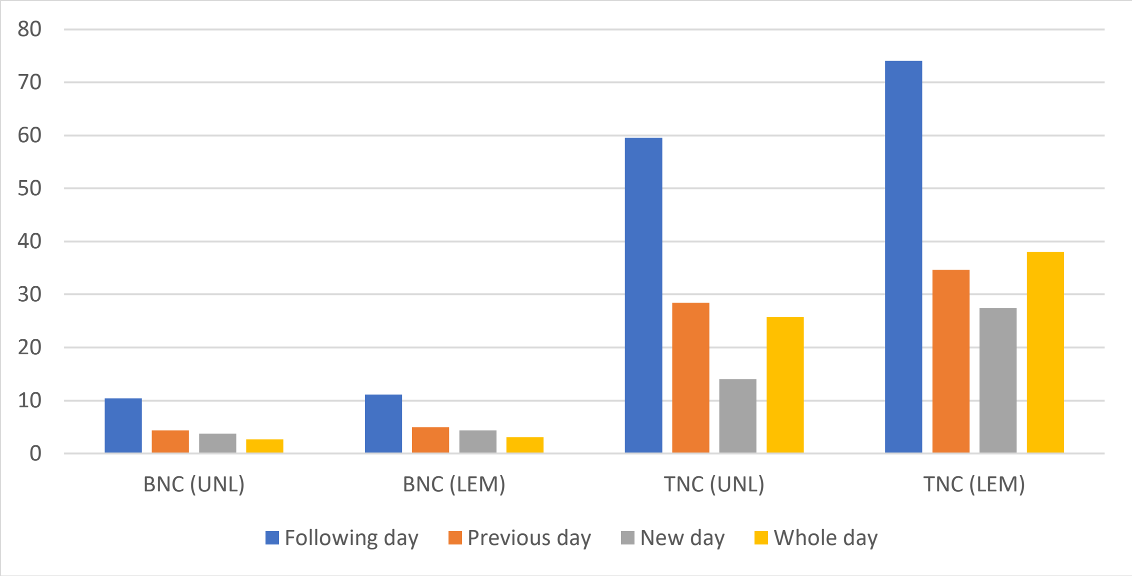

Figure 4.7 Relative collocation frequency scores (node-day)

As shown in Figure 4.8, the MI-scores revealed that all four unlemmatised combinations *ertesi gün*, *önceki gün*, *yeni gün*, and *bütün gün* obtain higher MI-scores of 6.52, 4.4, 0.48, 1.7 in Turkish than their English equivalents *following day*, *previous day*, *new day*, and *whole day* (3.4, 3.28, -0.3, 1.25 respectively). When they are lemmatised, all four combinations *ertesi gün*, *önceki gün*, *yeni gün*, and *bütün gün* reach higher MI-scores of 5.71, 3.58, 0.9, and 1.5 in Turkish than their English equivalents *following day*, *previous day*, *new day*, and *whole day* (2.88, 2.86, -0.69, 0.85 respectively). The LD-scores showed that all four unlemmatised combinations obtain higher LD-scores of 8.19, 7.1, 5.82, and 6.76 in Turkish than their English equivalents *following day*, *previous day*, *new day*, and *whole day* (6.44, 5.23, 4.63, 4.46 respectively). When they are lemmatised, all four combinations *Ertesi gün*, *Önceki gün*, *Yeni gün*, and *Bütün gün* reach higher LD-scores of 7.4, 6.31, 6.41, and 6.69 in Turkish than their English equivalents *following day*, *previous day*, *new day*, and *whole day* (5.95, 4.83, 4.37, and 4.09 respectively).

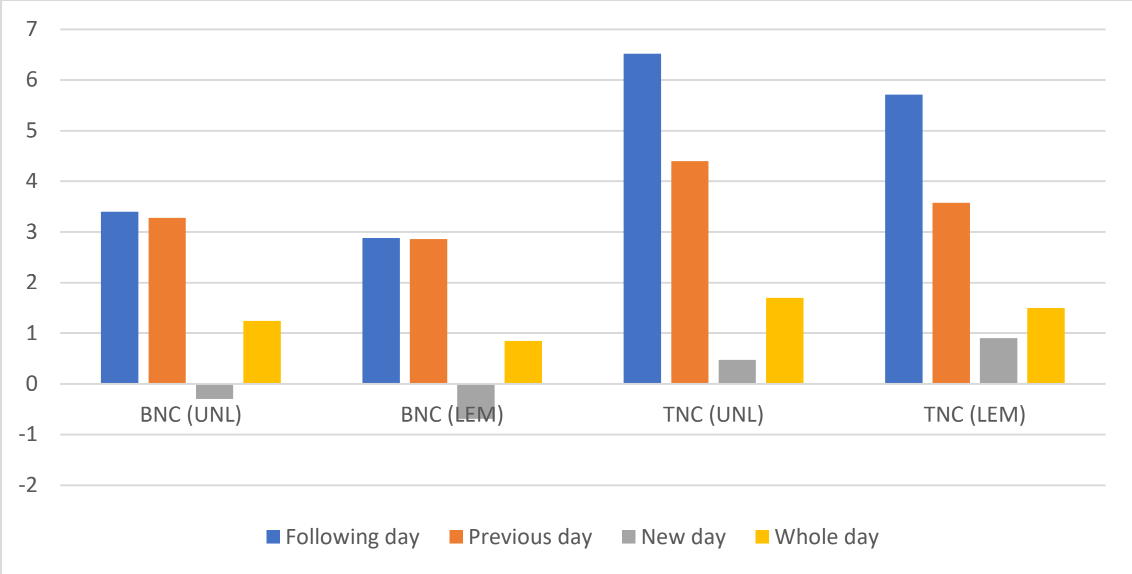

Figure 4.8 MI-scores (node-day)

The collocation *previous day* is also a notable example for the differences between lemmatised and unlemmatised RCF and association scores in Turkish and English. The unlemmatised collocation *previous day* (see *Figure 4.7*) obtains RCF score of 10.4, and when it lemmatised, it slightly increased to 11.09 in English. Its Turkish equivalent *önceki gün* has RCF score of 28.43, and when it is lemmatised, it increased to 34.66. The unlemmatised collocations *previous day* obtains LD-score of 5.23, and when it is lemmatised, it reaches LD-score of 4.83. Its Turkish equivalent collocation *önceki gün* obtains LD-score of 7.1, and when it is lemmatised, it has LD-score of 6.31. To understand the reason behind the large difference between the association scores of unlemmatised and lemmatised *previous day* in both languages, I explored the association counts of 5 case inflected (accusative, dative, genitive, locative, and ablative), and plural inflected forms of this collocation in English. As shown in Figure 4.9, the lemmatised collocations in both of the languages have lower association scores. The reason for this lower association counts for lemmatised collocation *previous day* should be that the case and plural inflected forms do not appear to be strongly associated adjective-noun combinations in Turkish, and the plural inflected form do not appear to be a strongly associated form in English (see Figure 4.9).

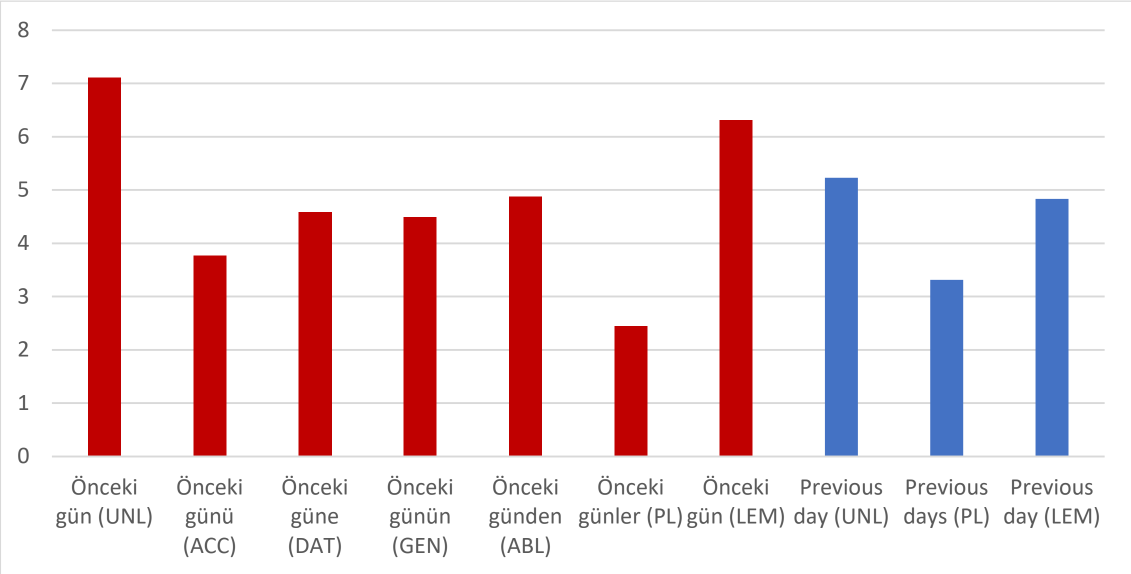

Figure 4.9 LD-scores for unlemmatised and lemmatised collocation *previous day*

Four adjective-noun combinations extracted through the node *way*, *only way* (Tek yol), *long way* (uzun yol), *other way* (diğer yol), and *different way* (farklı yol) were investigated for frequency and association counts in the BNC and TNC. As shown in Figure 4.10, the RCF scores showed that all four unlemmatised combinations occur at higher frequency in English than their Turkish equivalents. The unlemmatised combinations *only way, long way, other way,* and *different way* reach higher RCF scores of 19.63, 18.51, 18.27, and 6.67 in English than their Turkish equivalents *tek yol*, *uzun yol*, *diğer yol*, *farklı yol* (6.19, 5.91, 3.47, 2.8 respectively). When they are lemmatised, three out of four combinations *long way, other way,* and *different way* obtain higher RCF scores of 18.61, 25.98, and 20.32 in English than their Turkish equivalents *uzun yol*, *diğer yol*, *farklı yol* (13.79, 9.7, 7.02 respectively). The lemmatised combination *tek yol* reaches higher RCF score of 22.51 in Turkish than its English equivalent *only way* (19.77).

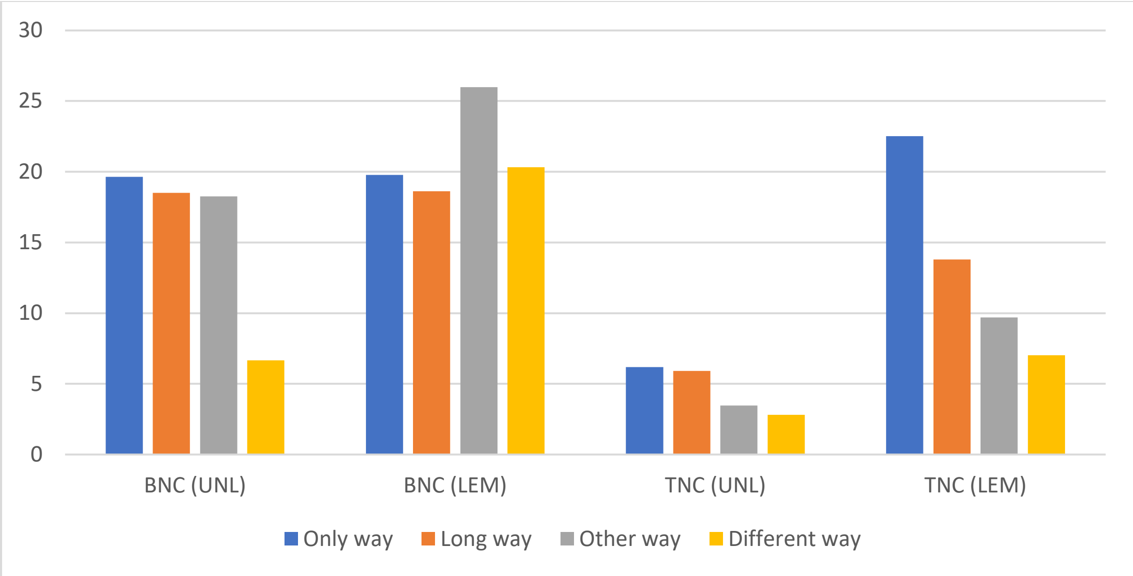

Figure 4.10 Relative collocation frequency scores (node-way)

As shown in Figure 4.11, the MI-scores showed that all four unlemmatised combinations *only way, long way, other way,* and *different way* reach higher MI-scores of 1.61, 2.49, 1, 1.24 in English than their Turkish equivalents *tek yol, uzun yol*, *diğer yol*, *farklı yol* (1.12, 1.44, 0.085, 0.53 respectively). When the combinations are lemmatised, three out of four combinations *long way, other way,* and *different way* obtain higher MI-scores of 2.29, 1.3, and 2.64 in English than their Turkish equivalents *uzun yol*, *diğer yol*, and *farklı yol* (0.97, -0.039, 0.17 respectively). The lemmatised collocation *tek yol* has higher MI-score of 1.29 than its English equivalent *only way* (0.96). The LD-scores revealed that all four unlemmatised combinations *only way, long way, other way,* and *different way* reach higher LD-scores of 6.45, 6.56, 6.34, and 5.11 in English than their Turkish equivalents *tek yol, uzun yol*, *diğer yol*, *farklı yol* (5.66, 5.68, 4.8, 4.63 respectively). When the combinations are lemmatised, all four combinations *only way, long way, other way,* and *different way* also reach higher LD-scores of 6.3, 6.38, 6.68, 6.53 in English than their Turkish equivalents *tek yol, uzun yol*, *diğer yol* and *farklı yol* (6.08, 5.4, 4.86, 4.44 respectively).

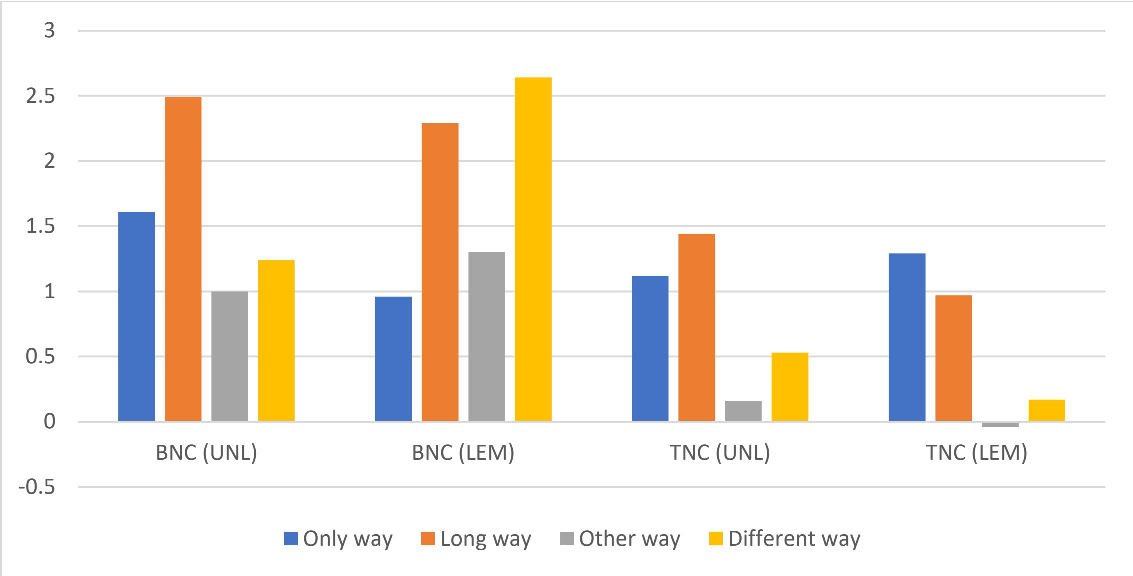

Figure 4.11 MI-scores (node-way)

Four adjective-noun combinations extracted through the node *man*, *young man* (genç adam), *old man* (yaşlı adam), *good man* (iyi adam), *big man* (büyük adam) were explored for frequency and association counts in the BNC and TNC. As shown in Figure 4.12, the RCF scores revealed that two out of four unlemmatised combinations occur at higher frequency in English than their Turkish equivalents. The combinations *young man* and *old man* reach higher RCF scores of 28.27, and 25.2 in English than their Turkish equivalents *genç adam* and *yaşlı adam* (16.79, and 14.68 respectively). The combinations *iyi adam* and *büyük adam* obtain higher RCF scores of 5.5, and 6.21 in Turkish than their English equivalents *good man* and *big man* (4.72, and 4.21 respectively). When the combinations are lemmatised, two out of four combinations *young man* and *old man* reach higher RCF scores of 41.22, and 28.86 in English than their Turkish equivalents *genç adam* and *yaşlı adam* (28 and 26.02 respectively). The lemmatised combinations *iyi adam* and *büyük adam* obtain higher RCF scores of 12.56, and 15.76 in Turkish than their English equivalents (6.61 and 5.01 respectively).

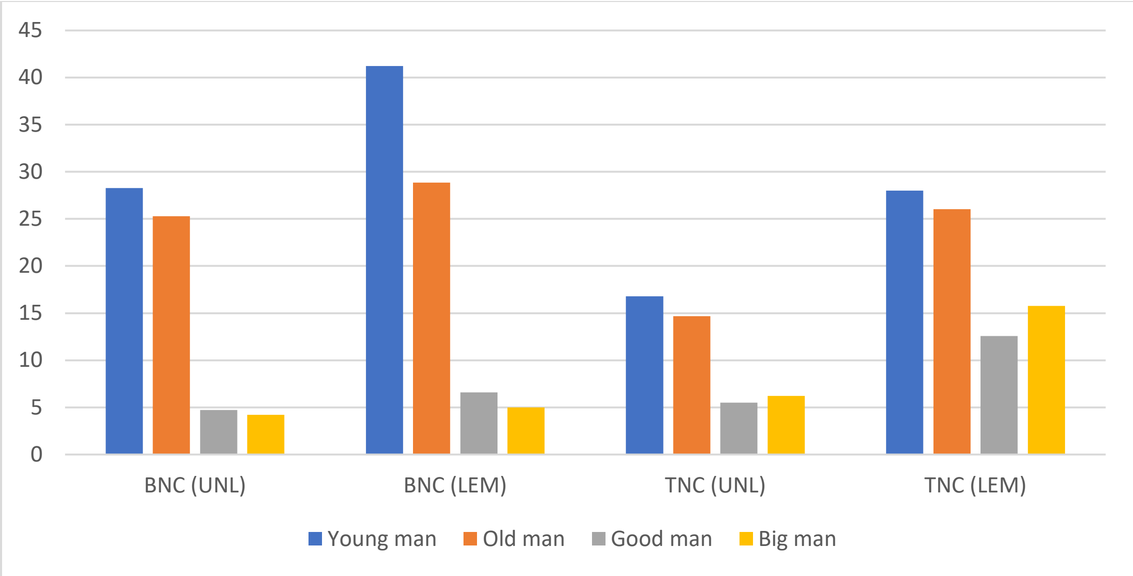

Figure 4.12 Relative collocation frequency scores (node-man)

As shown in Figure 4.13, the MI-scores showed that two out of four unlemmatised combinations *young man* and *big man* obtain higher MI-scores of 4.59, and 2.22 in English than their Turkish equivalents *genç adam* and *büyük adam* (3.91 and 0.76 respectively). The unlemmatised combinations *yaşlı adam* and *iyi adam* reach higher MI-scores of 5.14, and 1.21 in Turkish than their English equivalents *old man* and *good man* (3.72 and 0.68 respectively). When they are lemmatised, two out of four combinations *young man* and *big man* obtain higher MI-scores of 4.42, and 1.76 in English than their Turkish equivalents *genç adam* and *büyük adam* (3.39, and 0.48 respectively). The combinations *yaşlı adam* and *iyi adam* reach higher MI-scores of 4.72 and 0.59 in Turkish than their English equivalents *old man* and *good man* (3.2 and 0.46 respectively). Two out of four unlemmatised combinations *young man* and *old man* obtain higher LD-scores of 7.89 and 7.65 in English than their Turkish equivalents *genç adam* and *yaşlı adam* (7.48 and 7.41 respectively). The combinations *iyi adam* and *büyük adam* have higher LD-scores of 6.09 and 5.93 in Turkish than their English equivalents *good man* and *big man* (5.13 and 5.17 respectively). When they are lemmatised, two out of four combinations *young man* and *old man* reach higher LD-scores of 7.77 and 7.21 in English than their Turkish equivalents *genç adam* and *yaşlı adam* (7.07 and 7.02 respectively). The combinations *iyi adam* and *büyük adam* obtain higher LD-scores of 5.74 and 5.98 in Turkish than their English equivalents *good man* and *big man* (5.02 and 4.25 respectively).

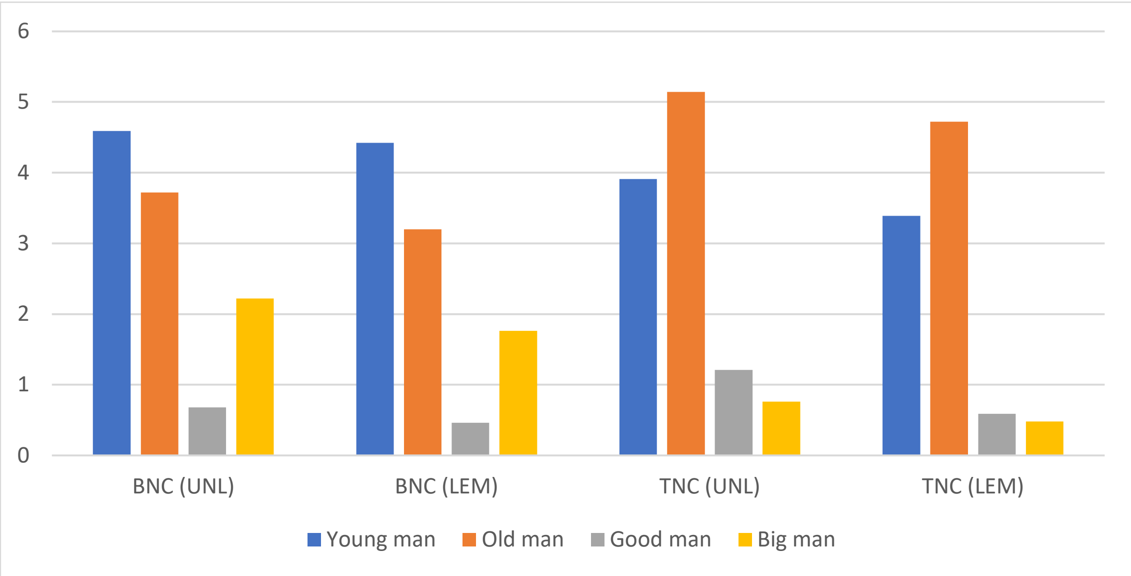

Figure 4.13 MI-scores (node-man)

In the mid-frequency band, 20 adjective-noun pairs were extracted from the BNC using five node words; *country, service, family, society, result*, and Turkish equivalents of these collocations were identified in the TNC.

Four adjective-noun combinations extracted through the node *country*, *other country* (diğer ülke), *whole country* (bütün ülke), *foreign country* (yabanci ülke), *new country* (yeni ülke) were investigated for frequency and association counts in the BNC and TNC. As shown in Figure 4.14, the RCF scores showed that all four unlemmatised combinations occur at higher frequency in Turkish than their English equivalents. The unlemmatised combinations *diğer ülke*, *bütün ülke*, *yabanci ülke*, and *yeni ülke* have higher RCF scores of 3.63, 2.36, 2.95, and 2.56 in Turkish than their English equivalents *other country*, *whole country*, *foreign country*, and *new country* (3.53, 2.24, 1.9, and 1.49 respectively). When they are lemmatised, all four combinations *diğer ülke*, *bütün ülke*, *yabanci ülke*, and *yeni ülke* obtain higher RCF scores of 50.41, 24.32, 24.62, and 18.35 in Turkish than their English equivalents *other country*, *whole country*, *foreign country*, and *new country* (27.24, 2.36, 3.29, and 2.18 respectively).

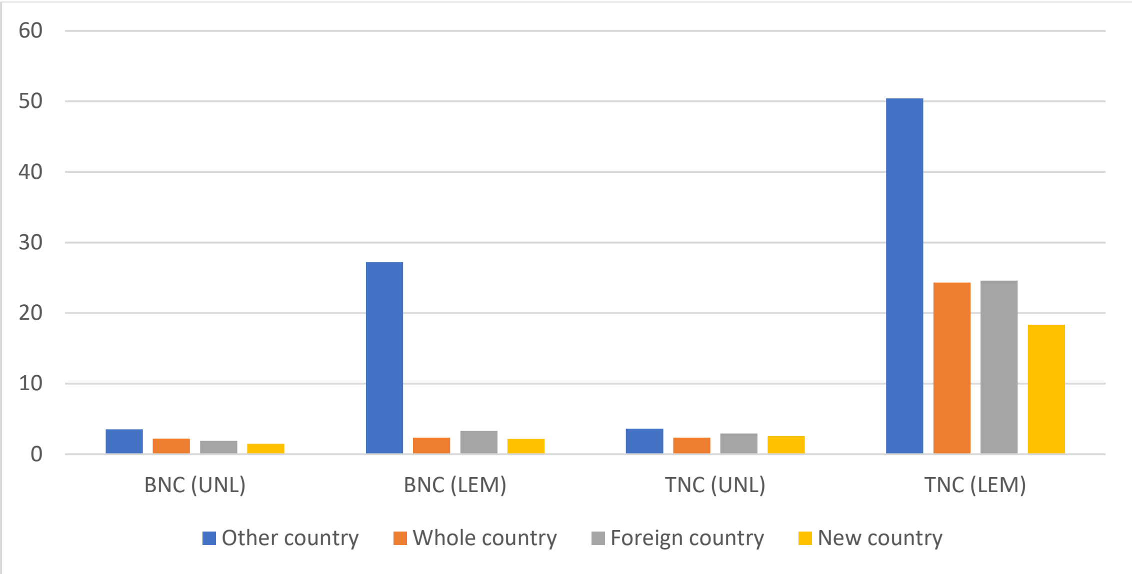

Figure 4.14 Relative collocation frequency scores (node-country)

As shown in Figure 4.15, the MI-scores revealed that three out of four unlemmatised combinations *diğer ülke*, *yabanci ülke*, and *yeni ülke* reach higher MI-scores of 1.73, 3.07, and 0.8 in Turkish than their English equivalents *other country*, *foreign country*, and *new country* (0.24, 2.61, and -0.68 respectively). The unlemmatised combination *whole country* obtains higher MI-score of 1.95 in English than its equivalent in Turkish *bütün ülke* (0.86). When they are lemmatised, two out of four combinations *other country* and *whole country* reach higher MI-scores of 2.58, and 1.41 in English than their Turkish equivalents *diğer ülke* and *bütün ülke* (2.26 and 0.94 respectively). The lemmatised combinations *yabanci ülke*, and *yeni ülke* obtain higher MI-scores of 2.95 and 0.19 in Turkish than their English equivalents *foreign country*, and *new country* (2.79 and -0.75 respectively). The LD-scores showed that all four unlemmatised combinations *diğer ülke*, *bütün ülke*, *yabanci ülke*, and *yeni ülke* have higher LD-scores of 6.06, 5.35, 6.13 and 5.47 in Turkish than their English equivalents *other country*, *whole country*, *foreign country*, and *new country* (5.05, 5.05, 4.91 and 3.94 respectively). When they are lemmatised, three out of four combinations *bütün ülke*, *yabanci ülke*, and *yeni ülke* have higher LD-scores of 6.09, 6.23, and 5.64 in Turkish than their English equivalents *whole country*, *foreign country*, and *new country* (4.58, 5.13 and 4.09 respectively). The lemmatised combination *other country* reaches higher LD-score of 7.63 in English than its Turkish equivalent *diğer ülke* (7.17).

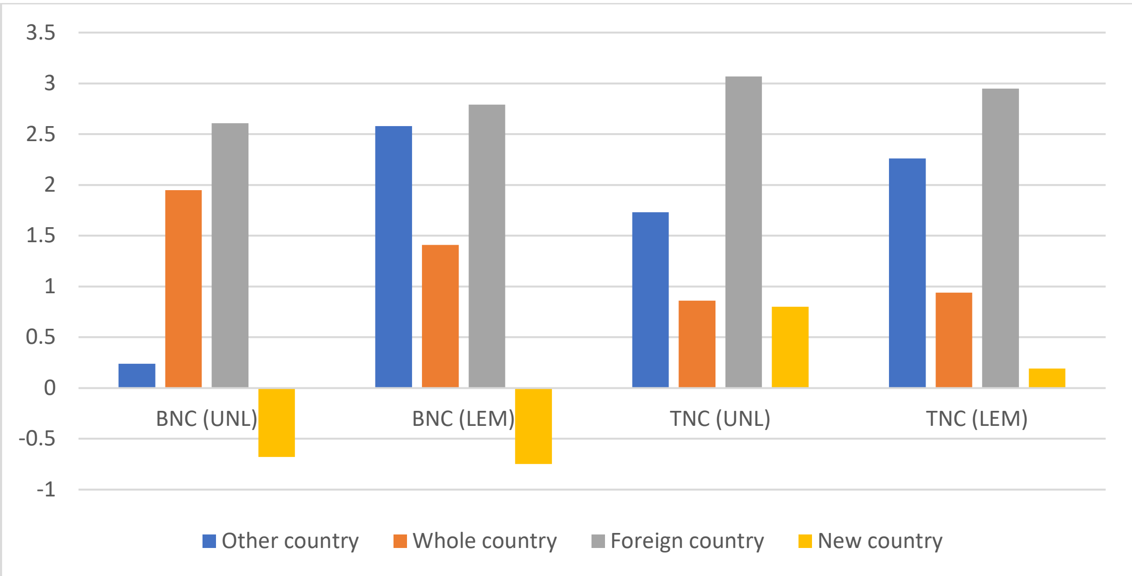

Figure 4.15 MI-scores (node-country)

Four adjective-noun combinations extracted through the node *family, whole family* (bütün aile), *other family* (*diğer* aile)*, large family* (büyük aile), *new family* (yeni aile) were investigated for frequency and association counts in the BNC and TNC. As shown in Figure 4.16, the RCF scores showed that two out of four unlemmatised combinations occur at higher frequency in English than their Turkish equivalents. The combinations *other family* and *new family* reach higher RCF scores of 2.27 and 1.92 in English than their Turkish equivalents *diğer aile* and *yeni aile* (2.21 and 1.49 respectively). The combinations *bütün aile* and *büyük aile* have higher RCF scores of 3.21 and 3.72 in Turkish than their English equivalents *whole family* and *large family* (3.1 and 2.13 respectively). When they are lemmatised, all four combinations *bütün aile*, *diğer aile*, *büyük aile* and *yeni aile* reach higher RCF scores of 8.97, 5.46, 10.3 and 4.69 in Turkish than their English equivalents *whole family*, *other family*, *large family*, and *new family* (3.53, 3.84, 3.37 and 2.54 respectively).

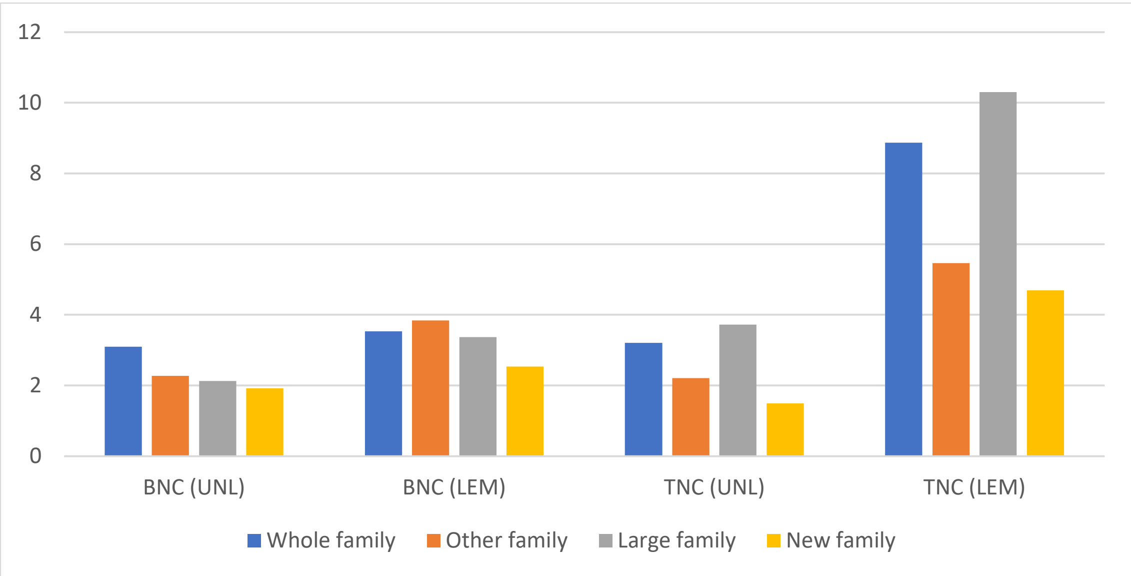

Figure 4.16 Relative collocation frequency scores (node-family)

As shown in Figure 4.17, the MI-scores showed that three out of four unlemmatised combinations *whole family*, *large family*, and *new family* obtain higher MI-scores of 2.31, 1.58, and -0.42 in English than their Turkish equivalents *bütün aile*, *büyük aile* and *yeni aile* (0.91, 0.58 and -0.53 respectively). The unlemmatised combination *diğer aile* reaches higher MI-scores of 0.63 in Turkish than its English equivalent *other family* (-0.49). When they are lemmatised, three out of four combinations *whole family*, *large family*, and *new family* have higher MI-scores of 2.18, 1.93, -0.33 in English than their Turkish equivalents *bütün aile*, *büyük aile* and *yeni aile* (0.86, 0.53, -0.4 respectively). The lemmatised combination *diğer aile* obtains higher MI-scores of 0.42 in Turkish than its English equivalent *other family* (-0.05). The LD-scores revealed that two out of four unlemmatised combinations *whole family*, and *new family* reach higher LD-scores of 5.43 and 4.25 in English than their Turkish equivalents *bütün aile* and *yeni aile* (5.32 and 4.06 respectively). The unlemmatised combinations *diğer aile* and *büyük aile* have higher LD-scores of 4.9 and 5.26 in Turkish than their English equivalents *other family* and *large family* (4.36 and 4.86 respectively). When they are lemmatised, all four combinations *bütün aile*, *diğer aile*, *büyük aile* and *yeni aile* reach LD-scores of 5.79, 5.14, 5.86 and 4.78 in Turkish than their English equivalents *whole family*, *other family*, *large family*, and *new family* (5.34, 4.93, 5.25 and 4.45 respectively).

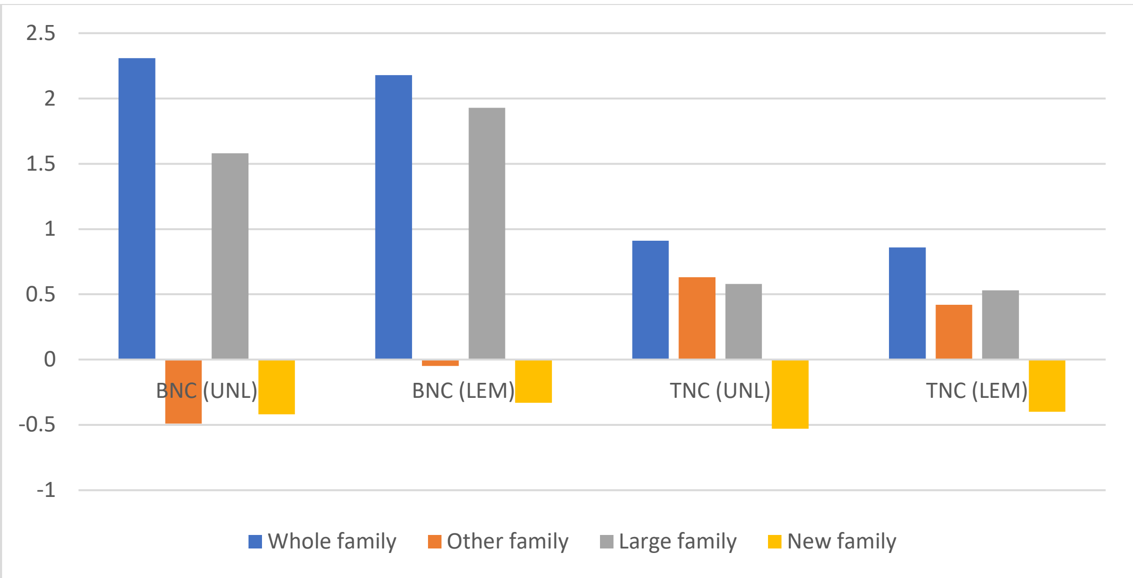

Figure 4.17 MI-scores (node-family)

Four adjective-noun combinations extracted through the node *society*, *civil society* (sivil toplum), *modern society* (modern toplum), *new society* (yeni toplum), *capitalist society* (kapitalist toplum) were explored for frequency and association counts in the BNC and TNC. As shown in Figure 4.18, the RCF scores revealed that three out of four unlemmatised combinations occur at higher frequency in Turkish than their English equivalents. The combinations *sivil toplum*, *modern toplum*, and *yeni toplum*, have higher RCF scores of 45.54, 3.11 and 3.39 in Turkish than their English equivalents *civil society*, *modern society*, and *new society* (45.54, 3.11, 3,39 respectively). The combination *capitalist society* obtains a higher RCF score of 1.55 in English than its Turkish equivalent *kapitalist toplum* (0.8). When they are lemmatised, all four combinations *sivil toplum*, *modern toplum*, *yeni toplum*, and *kapitalist toplum* reach higher RCF scores of 59.13, 9.45, 9.96, and 5.05 in Turkish than their English equivalents *civil society*, *modern society*, *new society*, and *capitalist society* (3, 3.97, 2.59 and 2.31 respectively).

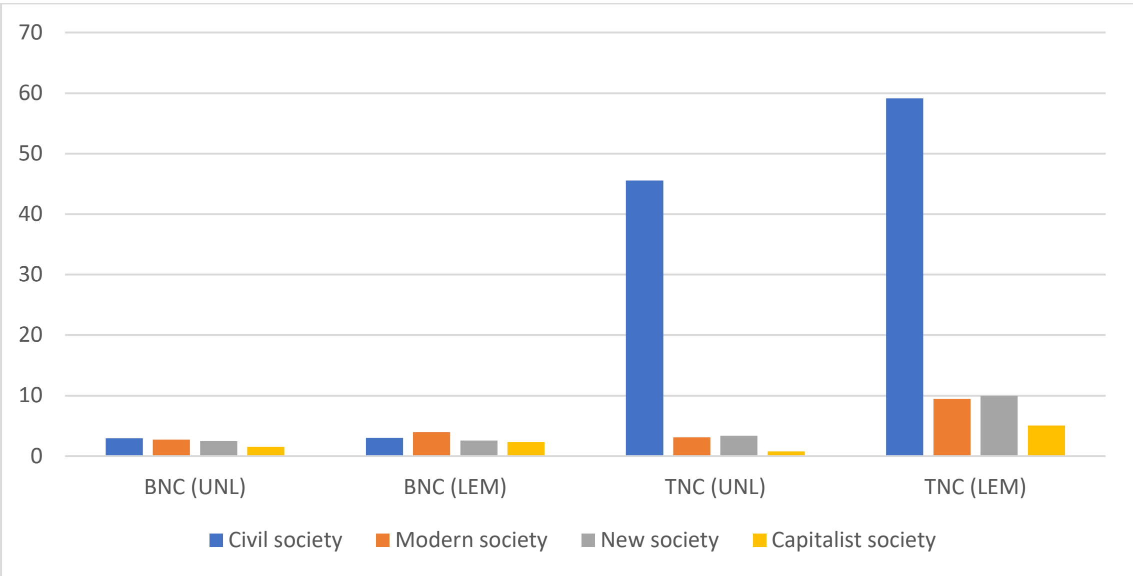

Figure 4.18 Relative collocation frequency scores (node-society)

As shown in Figure 4.19, the MI-scores showed that two out of four unlemmatised combinations *sivil toplum*, and *yeni toplum* have higher MI-scores of 8.03 and 0.67 in Turkish than their English equivalents *civil society*, and *new society* (4.85 and 0.45 respectively). The unlemmatised combinations *modern society* and *capitalist society* in English obtain higher MI-scores of 3.88 and 5.5 than their Turkish equivalents *modern toplum*, and *yeni toplum* (3.35 and 3.51 respectively). When they are lemmatised, three out of four combinations *sivil toplum*, *modern toplum*, and *yeni toplum* reach higher MI-scores of 6.57, 3.8 and 0.38 in Turkish than their English equivalents *civil society*, *modern society*, and *new society* (3.4, 3.2 and -0.67 respectively). The lemmatised combination *capitalist society* obtains a higher MI-score of 4.97 in English than its Turkish equivalent *kapitalist toplum* (4.32). The LD-scores revealed that two out of four unlemmatised combinations *modern society* and *capitalist society* have higher LD-scores of 5.85 and 5.13 in English than their Turkish equivalents *modern toplum* and *kapitalist toplum* (5.37 and 4.2 respectively). The unlemmatised combinations *sivil toplum* and *yeni toplum* obtain higher LD-scores of 9.94, and 5.25 in Turkish than their English equivalents *civil society*, and *new society* (5.99 and 4.9 respectively). When they are lemmatised, the combinations *modern society* and *capitalist society* reach higher LD-scores of 6.16 and 5.48 in English than their Turkish equivalents *modern toplum* and *kapitalist toplum* (5.92 and 5.04 respectively). The lemmatised combinations *sivil toplum* and *yeni toplum* have higher LD-scores of 8.57 and 5.65 in Turkish than their English equivalents *civil society*, and *new society* (5.79 and 4.85 respectively).

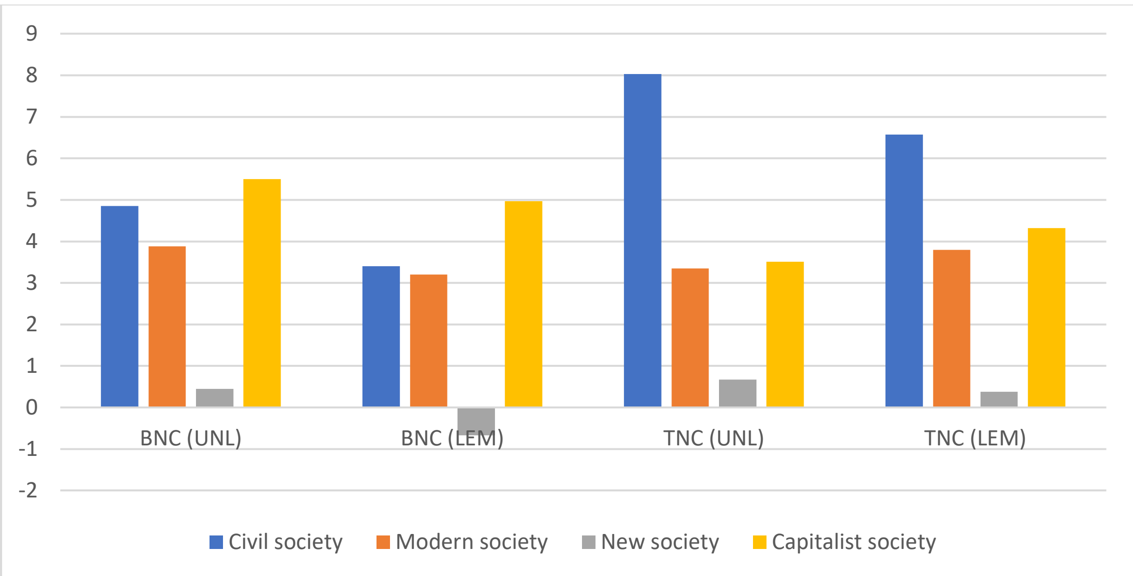

Figure 4.19 MI-scores (node-society)

The collocation *civil society* is another example for the differences between lemmatised and unlemmatised RCF and association scores in Turkish and English. The unlemmatised collocation *civil society* reaches RCF score of 2.95, and when it is lemmatised, it slightly increased to 3 in English. Its Turkish equivalent *sivil toplum* obtains RCF score of 45.54, and when it lemmatised, it increased to 59.13 in Turkish. The unlemmatised collocation *civil society* reaches LD-score of 5.99 in English, and when it lemmatised, it slightly decreased to 5.79. Its Turkish equivalent *sivil toplum* obtains 9.94, and when it lemmatised, it decreased to 8.57. To understand the reason behind the differences between the association scores of unlemmatised and lemmatised collocation *civil society*, I explored the association counts of 5 case inflected (accusative, dative, genitive, locative, and ablative), and plural inflected forms of *sivil toplum* in Turkish, and plural inflected forms of this collocation in English (see Figure 4.20). The reason for lower association counts for lemmatised collocations *civil society* in English, and *sivil toplum* in Turkish should be that the case inflected and plural forms in Turkish and the plural inflected form in English appear to be weaker associated combinations than the base form collocations in both languages

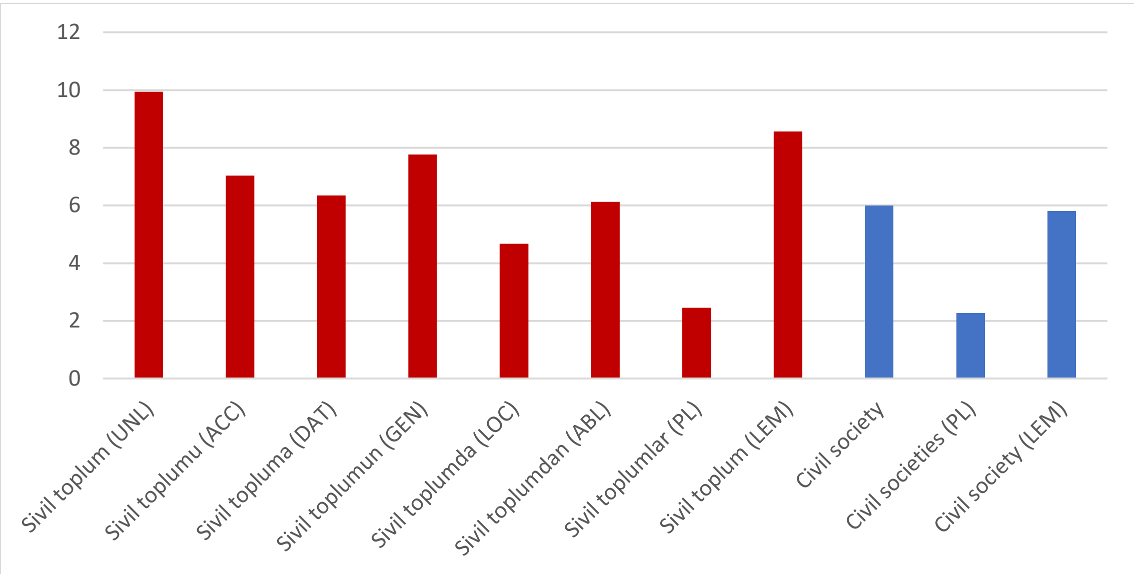

Figure 4.20 LD-scores for unlemmatised and lemmatised collocation *civil society*

Four adjective-noun combinations extracted through the node *result*, *likely result* (olası sonuç), *new result* (yeni sonuç), *good result* (iyi sonuç), and *different result* (farklı sonuç) were investigated for frequency and association counts in the BNC and TNC. As shown in Figure 4.21, the RCF scores revealed that two out of four unlemmatised combinations occur at higher frequency in English than their Turkish equivalents. The combinations *likely result* and *new result* have higher RCF scores of 1.31 and 1.14 in English than their Turkish equivalents *olası sonuç* and *yeni sonuç* (0.15 and 0.94 respectively)**.** The unlemmatised combinations *iyi sonuç* and *farklı sonuç* obtain higher RCF scores of 5.89 and 1.06 in Turkish than their English equivalents *good result* and *different result* (1.08 and 0.88 respectively). When they are lemmatised, all four combinations *olası sonuç*, *yeni sonuç*, *iyi sonuç*, and *farklı sonuç* reach higher RCF scores of 2.19, 8.85, 15.66, and 9.11 in Turkish than their English equivalents *likely result***,** *new result*, *good result*, and *different result* (1.59, 2.29, 3.63, and 2.06 respectively).

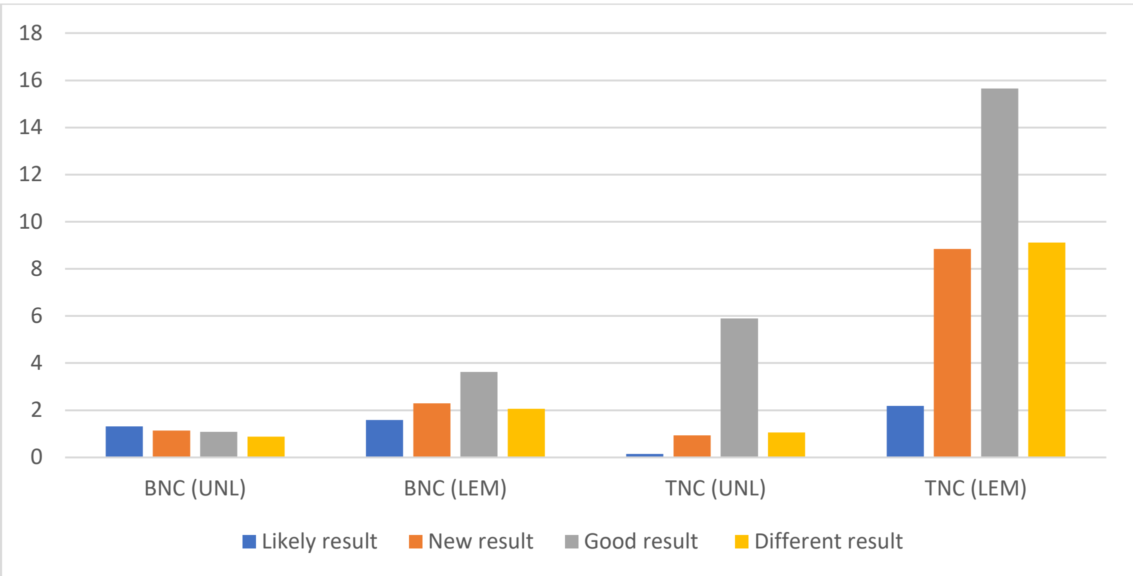

Figure 4.21 Relative collocation frequency scores (node-result)

As shown in Figure 4.22, the MI-scores revealed that two out of four unlemmatised combinations *likely result*, and *new result* have higher MI-scores of 2.08 and -0.55 in English than their Turkish equivalents *olası sonuç*, and *yeni sonuç* (1.18 and -0.92 respectively). The unlemmatised combinations *iyi sonuç* and *farklı sonuç* obtain higher MI-scores of 1.95 and 0.53 in Turkish than their English equivalents *good result* and *different result* (-0.013 and 0.45 respectively). When they are lemmatised, all four combinations *olası sonuç*, *yeni sonuç*, *iyi sonuç*, and *farklı sonuç* reach higher MI-scores of 2.46, -0.21, 0.84, and 1.11 in Turkish than their English equivalents *likely result*, *new result*, *good result*, and *different result* (1.42, -0.48, 0.79, 0.75 respectively). The LD-scores showed that two out of four unlemmatised combinations *likely result*, and *new result* reach higher LD-scores of 4.78 and 3.85 in English than their Turkish equivalents *olası sonuç*, and *yeni sonuç* (2.06 and 3.52 respectively). The unlemmatised combinations *iyi sonuç*, and *farklı sonuç* have higher LD-scores of 6.28 and 4.27 in Turkish than their English equivalents *good result* and *different result* (4.04 and 3.99 respectively). When they are lemmatised, one combination *likely result* reaches higher LD-score of 4.22 in English than its Turkish equivalent *olası sonuç* (3.4). Three lemmatised combinations *yeni sonuç*, *iyi sonuç*, and *farklı sonuç* obtain higher LD-scores of 5.13, 6, and 5.34 in Turkish than their English equivalents *new result*, *good result*, and *different result* (4.3, 5.14, and 4.48 respectively).

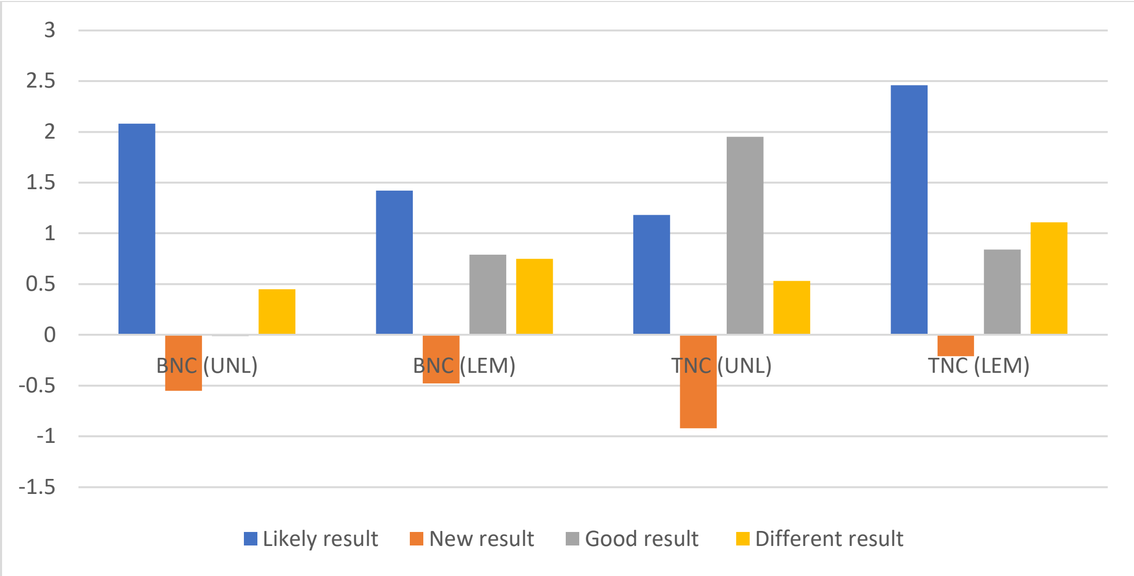

Figure 4.22 MI-scores (node-result)

The collocation *likely result* is also a notable example for the differences between lemmatised and unlemmatised RCF and association scores in Turkish and English. The unlemmatised collocation *likely result* obtains RCF score of 1.41 in English, and when it is lemmatised, it slightly increased to 1.59. Its Turkish equivalent unlemmatised *olası sonuç* reaches RCF score of 0.15 and when it is lemmatised, it increased to 2.19. The unlemmatised collocation *likely result* obtains LD-score of 4.78 and when it lemmatised, it slightly decreased to 4.22 in English. Its Turkish equivalent *olası sonuç* reaches LD-score of 2.06 in English, and when it is lemmatised, it increased to 3.4. To understand the reason behind the differences between the association scores of unlemmatised and lemmatised collocation *likely result* in English and Turkish, association counts of 5 case inflected (accusative, dative, genitive, locative, and ablative), and plural inflected forms of *olası sonuç* in Turkish, and plural inflected forms of its English equivalent *likely result* were explored. The reason for lower association count for lemmatised collocation *likely result* in English is that plural inflected form of this collocation is weakly-associated (see Figure 4.23). The reason for higher association count for *olası sonuç* in Turkish is that dative and genitive inflected forms which have higher LD-scores of 4.32 and 4.1 than the base form of the collocation (2.06).

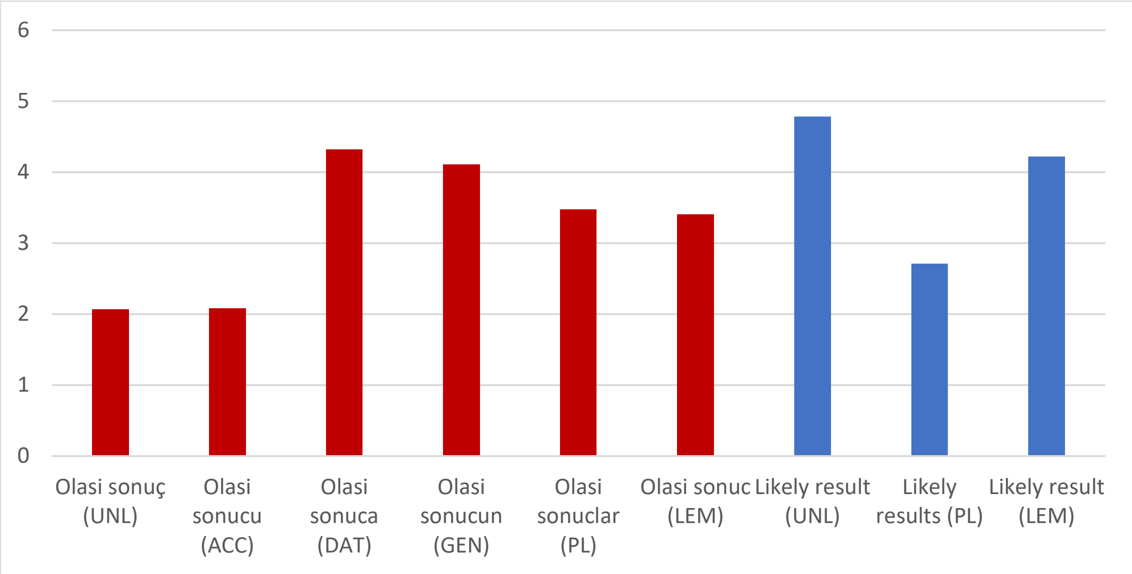

Figure 4.23 LD-scores for unlemmatised and lemmatised collocation Likely result

Four adjective-noun combinations extracted through the node *industry, nuclear industry* (nükleer sanayi), *new industry* (yeni sanayi)*, local industry* (yerel sanayi), *and heavy industry* (ağır sanayi) were explored for frequency and association counts in the BNC and TNC. As shown in Figure 4.24, the RCF scores showed that three out of four unlemmatised combinations occur at higher frequency in English than their Turkish equivalents. The combinations *nuclear industry*, *local industry*, and *heavy industry* reach higher RCF scores of 2.88, 1.49, and 1.16 in English than their Turkish equivalents *nükleer sanayi*, *yerel sanayi*, and *ağır sanayi* (0.13, 0.31 and 0.88 respectively). The unlemmatised combination *yeni sanayi* obtains higher RCF score of 2.01 in Turkish than its English equivalent *new industry* (1.89). When they are lemmatised, three out of four combinations *nuclear industry*, *local industry*, and *heavy industry* have higher RCF scores of 2.97, 2.09 and 1.83 in English than their Turkish equivalents *nükleer sanayi*, *yerel sanayi*, and *ağır sanayi* (0.33, 0.47, and 1.67 respectively). The combination *yeni sanayi* obtains higher RCF score of 3.8 in Turkish than its English equivalent *New industry* (3.59).

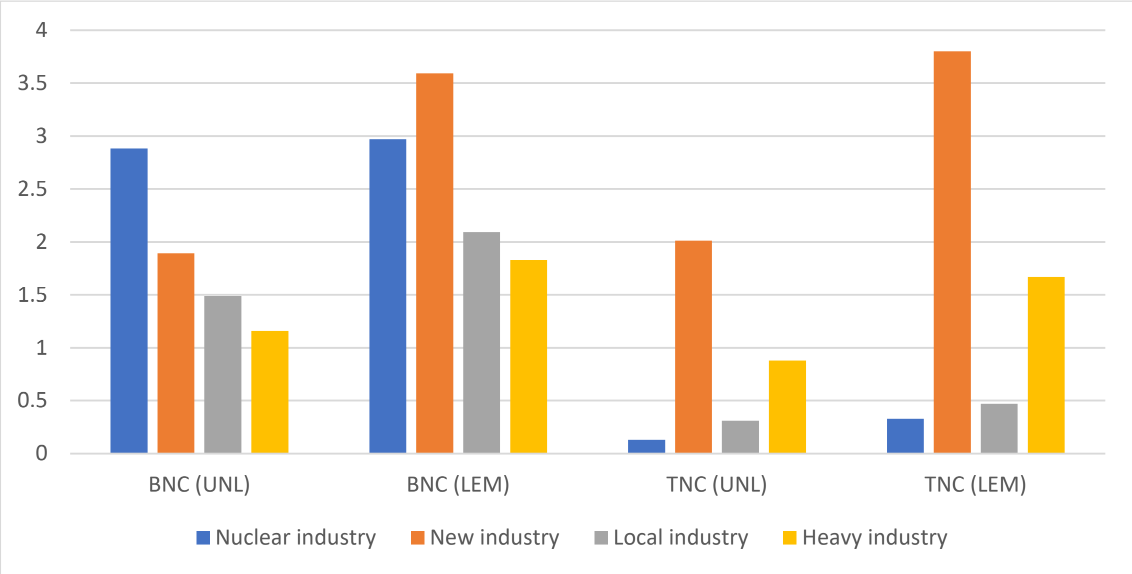

Figure 4.24 Relative collocation frequency scores (node-industry)

As shown in Figure 4.25, the MI-scores showed that three out of four unlemmatised combinations *nuclear industry*, *local industry*, and *heavy industry* obtain higher MI-scores of 4.86, 1.41, and 3.39 in English than their Turkish equivalents *nükleer sanayi*, *yerel sanayi*, and *ağır sanayi* (0.68, 0.85, 1.69 respectively). The unlemmatised combination *yeni sanayi* reaches a higher MI-score of 0.35 in Turkish than its English equivalent *new industry* (0.32). When they are lemmatised, two out of four combinations *nuclear industry*, and *heavy industry* have higher MI-scores of 3.66, and 1.26 in English than their Turkish equivalents *nükleer sanayi*, and *ağır sanayi* (1.26 and 1.91 respectively). The lemmatised combinations *yeni sanayi* and *yerel sanayi* obtain higher MI-scores of 0.57 and 0.74 in Turkish than their English equivalents *new industry* and *local industry* (0.003 and 0.65 respectively). The LD-scores showed that three out of four unlemmatised combinations *nuclear industry*, *local industry*, and *heavy industry* reach higher LD-scores of 6.2, 4.87, and 4.88 than its Turkish equivalents *nükleer sanayi*, *yerel* *sanayi*, and *ağır sanayi* (2.02, 3.11 and 4.49 respectively). The unlemmatised combination *yeni sanayi* obtains an LD-score of 4.68 in Turkish than its English equivalent *new industry* (4.65). When they are lemmatised, all four combinations *nuclear industry*, *new industry*, *local industry*, and *heavy industry* have higher LD-scores of 5.97, 5.43, 5.14, and 5.27 in English than their Turkish equivalents *nükleer sanayi*, *yeni sanayi, yerel sanayi* and *ağır sanayi* (2.65, 5.28, 3.07, and 4.82 respectively).

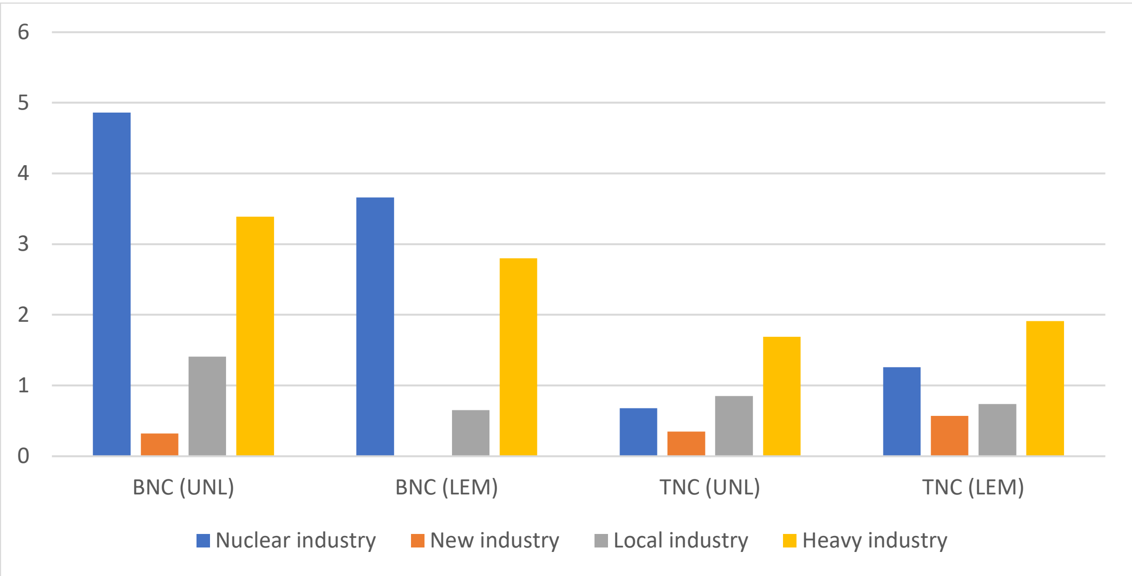

Figure 4.25 MI-scores (node-industry)

Tables 4.5 4.6, 4.7 4.8 demonstrate the unlemmatised and lemmatised adjective-noun combinations' RCF, LD, and Delta P counts in high- and mid-frequency bands in both languages. Therefore, it is possible to see the differences of RCF and association counts of lemmatised and unlemmatised word combinations in both languages. Table 4.5Table 4.6 show the unlemmatised and lemmatised adjective-noun combinations' RCF scores in both languages, and Tables Table 4.7Table 4.8 display the unlemmatised and lemmatised combinations' LD and Delta P scores.

In the high-frequency band, 11 out of 20 (55%) Turkish unlemmatised adjective-noun combinations, and 15 out of 20 (75%) Turkish lemmatised combinations occur at higher RCF scores than their English equivalents (see Table 4.5). The English unlemmatised adjective-noun combinations reach a mean RCF score of 12.64 (SD=10.97) with high score of 46.82 (*long time*), and low score of 2.65 (*whole day*). Turkish unlemmatised adjective-noun combinations reach a mean RCF score of 14.97 (SD=13.38) with high score of 59.59 (*following day*), and low score of 2.2 (*real world*). When they are lemmatised, English adjective-noun combinations reach a mean RCF score of 15.04 (SD=12.46) with high score of 47.19 (*long time*) and low score of 3.08 (*whole day*). Turkish lemmatised adjective-noun combinations have a mean RCF score of 29.75 (SD=18.25) with high score of 74.03 (*following day*), and 7.02 (*different way*). In the mid-frequency band, 11 out of 20 (55%) Turkish unlemmatised adjective-noun combinations, and 15 out of 20 (75%) Turkish lemmatised combinations occur at higher RCF scores than their English equivalents (see Table 4.6). The English unlemmatised adjective-noun combinations reach a mean RCF score of 2.19 (SD=0.96), with high score of 4.88 (*heavy industry*), and low score of 0.88 (*different result*). Turkish unlemmatised combinations have a mean RCF score of 5.67 (11.1), with high score of 45.54 (*civil society*), and low score of 0.13 (*nuclear industry*). When they are lemmatised, English adjective-noun combinations reach a mean RCF score of 3.02 (SD=0.86), with high score of 5.27 (*heavy industry*), and low score of 1.59 (*likely result*). Turkish lemmatised adjective-noun combinations have a mean RCF score of 13.79 (SD=15.67), with high score of 59.13 (*civil society*), and low score of 0.33 (*nuclear industry*).

Table 4.5 Relative collocation frequency scores in the high-frequency band

Adj-noun combinations	Unlemmatised RCF scores (BNC)	Unlemmatised RCF scores (TNC)	Lemmatised RCF scores (BNC)	Lemmatised RCF scores (TNC)
Short time (*Kısa zaman*)	13.82	11.4	14.09	49.25
Long time (Uzun zaman)	46.82	27.03	47.19	54.63
Good time (Iyi zaman)	11.24	15.09	13.75	26.24
Right time *(*Doğru *zaman)*	6.69	8.97	6.88	19.55
New world (Yeni dünya)	8.32	23.04	8.5	39.72
Real world (Gerçek dünya)	7.83	2.2	7.85	9.47
Outside world (Dış dünya)	6.35	7.3	6.38	26.91
Whole world (Bütün dünya)	5.08	16.25	5.11	59.78
Following day (Ertesi gün)	10.4	59.59	11.09	74.03
Previous day (Önceki gün)	4.34	28.43	4.97	34.66
New day (Yeni gün)	3.72	14	4.34	27.46
Whole day (Bütün gün)	2.65	25.79	3.08	38.08
Only way (Tek yol)	19.63	6.19	19.77	22.51
Long way (Uzun yol)	18.51	5.91	18.61	13.79
Other way (Di*ğer yol)*	18.27	3.47	25.98	9.7
Different way (Farkl*ı yol)*	6.67	2.8	20.32	7.02
Young man (Genç adam)	28.27	16.79	41.22	28
Old man (Yaşlı adam)	25.28	14.68	28.86	26.02
Good man (Iyi adam)	4.72	5.5	6.61	12.56
Big man (Büyük adam)	4.21	5.01	6.21	15.76

Table 4.6 Relative collocation frequency scores in the mid-frequency band

Adj-noun combinations	Unlemmatised RCF scores (BNC)	Unlemmatised RCF scores (TNC)	Lemmatised RCF scores (BNC)	Lemmatised RCF scores (TNC)
Other country (Diğer *ülke)*	3.53	27.24	3.63	50.41
Whole country (Bütün ülke)	2.24	2.36	2.36	24.32
Foreign country (Yabancı ülke)	1.9	2.95	3.29	24.62
New country (Yeni ülke)	1.49	2.18	2.56	18.35
Whole family (Bütün aile)	3.1	3.53	3.53	8.87
Other family (Di*ğer aile)*	2.27	2.21	3.84	5.46
Large family (Büyük aile)	2.13	3.72	3.37	10.3
New family (Yeni aile)	1.92	1.49	2.54	4.69
Civil society (Sivil toplum)	2.95	45.54	3	59.13
Modern society (Modern topl.)	2.75	3.11	3.97	9.45
New society (Yeni toplum)	2.46	3.39	2.59	9.96
Capitalist society (Kapitalist topl.)	1.55	0.8	2.31	5.05
Likely result (Olası sonuç)	1.31	0.15	1.59	2.19
New result (Yeni sonuç)	1.14	0.94	2.29	8.85
Good result (Iyi sonuç)	1.08	5.89	3.63	15.66
Different result (Farklı sonuç)	0.88	1.06	2.06	9.11
Nuclear industry (Nükleer sanayi)	2.88	0.13	2.97	0.33
New industry (Yeni sanayi)	1.89	2.01	3.59	3.8
Local industry (Yerel sanayi)	1.49	0.31	2.09	0.47
Heavy industry (Ağır sanayi)	4.88	4.49	5.27	4.82

As can be seen in Table 4.7, 11 out of 20 (55%) Turkish unlemmatised adjective-noun combinations and 12 out of 20 (60%) Turkish lemmatised combinations reach higher LD-scores than their English equivalents in the high-frequency band. The English unlemmatised adjective-noun collocations have a mean LD-score of 5.83 (SD=1) with high score of 7.89 (*young man*) and low score of 4.42 (*right time*). The Turkish unlemmatised adjective-noun combinations obtain a mean LD-score of 6.07 (SD=1.06) with high score of 8.19 (*following day*), and low score of 4.42 (*real world*). When they are lemmatised, English adjective-noun combinations reach a mean LD-score of 5.72 (SD=1.05) with high score of 7.77 (*young man*), and low score of 4.09 (*whole day*). Turkish lemmatised adjective-noun combinations obtain a mean LD-score of 6.15 (SD=0.87) with high score of 7.48 (*whole world*), and low score of 4.44 (*different way*). As can also be seen in Table 4.7, 16 out of 20 (80%) lemmatised combinations in English and 7 out of 20 (35%) lemmatised combinations in Turkish have lower LD-scores than their unlemmatised forms. Furthermore, the vast majority of the Delta P scores for both English and Turkish combinations show that the directions of the association appear to be from adjectives to nouns in both languages.

As can be seen in Table 4.8, 11 out of 20 (55%) Turkish unlemmatised adjective-noun combinations and 12 out of 20 (60%) Turkish lemmatised combinations reach higher LD-scores than their English equivalents in the mid-frequency band. The English unlemmatised combinations have a mean LD-score of 4.84 (SD=0.66) with high score of 6.2 (*nuclear industry*), and low score of 3.85 (*new result*). The Turkish unlemmatised combinations have a mean LD-score of 4.88 (SD=1.69) with high score of 9.94 (*civil society*), and low score of 2.02 (*nuclear industry*). When they are lemmatised, English adjective-noun combinations reach a mean a LD-score of 5.18 (SD=0.81) with high score of 7.63 (*other country*), and low score of 4.09 (*new country*). Turkish lemmatised adjective-noun combinations reach a mean LD-score of 5.37 (SD=1.32) with high score of 8.57 (*civil society*), and low score of 3.07 (*local industry*). Table 4.8 also shows that 6 out of 20 (30%) lemmatised combinations in English and 3 out of 20 (15%) lemmatised combinations in Turkish have lower LD-scores than their unlemmatised forms. Similar to the results of high-frequency bands combinations, the vast majority of the

Delta P scores for both English and Turkish combinations show that the directions of the association appear to be from adjectives to nouns in both languages.

Table 4.7 Association statistics in the high-frequency band

Adj-noun	Unl Log Dice (BNC)	Unl Log Dice (TNC)	Lem Log Dice (BNC)	Lem Log Dice (TNC)	Delta P for adj. (BNC)	Delta P for noun (BNC)	Delta P for adj. (TNC)	Delta P for noun (TNC)
Short time (*Kısa zaman*)	5.57	5.14	5.35	6.56	.060	.001	.016	.0007
Long time (Uzun zaman)	7.28	6.35	7.05	6.69	.073	.004	.028	.0019
Good time (Iyi zaman)	5.18	5.43	5.24	5.58	.004	.0003	.001	.0001
Right time (*Doğru zaman)*	4.42	4.73	4.23	5.17	-.001	-.0001	-.000	-.000
New world (Yeni dünya)	5.84	7.19	5.85	6.85	.003	.001	.011	.004
Real world (Gerçek dünya)	6.1	4.42	6.08	4.93	.030	.002	.003	.0003
Outside world (Dış dünya)	5.81	5.91	5.79	6.45	.026	.001	.018	.001
Whole world (Bütün dünya)	5.45	6.78	5.43	7.48	.013	.001	.003	.009
Following day (Ertesi gün)	6.44	8.19	5.95	7.4	.035	.002	.595	.009
Previous day (Önceki gün)	5.23	7.1	4.83	6.31	.031	.001	.132	.004
New day (Yeni gün)	4.63	5.82	4.37	6.41	-.0006	-.0002	.002	.0006
Whole day (Bütün gün)	4.46	6.76	4.09	6.69	.005	.0004	.014	.002
Only way (Tek yol)	6.45	5.66	6.3	6.08	.007	.001	.003	.001
Long way (Uzun yol)	6.56	5.68	6.38	5.4	.026	.002	.005	.001
Other way (Di*ğer yol)*	6.34	4.8	6.68	4.86	.005	.001	.0003	.0001
Different way (Farklı *yol)*	5.11	4.63	6.53	4.44	.008	.0006	.001	.0002
Young man (Genç adam)	7.89	7.48	7.77	7.07	.082	.007	.037	.005
Old man (Yaşlı adam)	7.65	7.41	7.21	7.02	.043	.006	.092	.004
Good man (Iyi adam)	5.13	6.09	5.02	5.74	.002	.0005	.001	.0005
Big man (Büyük adam)	5.1	5.93	4.25	5.98	.013	.0009	.0005	0003

Table 4.8 Association statistics in the mid-frequency band

Adj-noun	Unl Log Dice (BNC)	Unl Log Dice (TNC)	Lem Log Dice (BNC)	Lem Log Dice (TNC)	Delta P for adj. (BNC)	Delta P for noun (BNC)	Delta P for adj. (TNC)	Delta P for noun (TNC)
Other country (Diğer ülke)	5.05	6.06	7.63	7.17	.0003	.0002	.002	.001
Whole country (Bütün ülke)	5.05	5.35	4.58	6.09	.0054	.0008	.0003	.0002
Foreign country (Yabancı ülke)	4.91	6.13	5.13	6.23	.0098	.0008	.009	.001
New country (Yeni ülke)	3.94	5.47	4.09	5.64	-.0004	-0007	.0000	.0000
Whole family (Bütün aile)	5.43	5.32	5.34	5.79	.0012	.0008	.0012	.0010
Other family (Diğer aile)	4.36	4.9	4.93	5.14	-.0005	-.0004	.0007	.0005
Large family (Büyük aile)	4.86	5.26	5.25	5.86	.0040	.0006	.0007	.0008
New family (Yeni aile)	4.25	4.06	4.45	4.78	-.0005	-.0003	-.0004	-.0004
Civil society (Sivil toplum)	5.99	9.94	5.79	8.57	.032	.0019	.36	.032
Modern society (Modern toplum)	5.85	5.37	6.16	5.92	.019	.001	.012	.001
New society (Yeni toplum)	4.9	5.25	4.85	5.65	.0005	.0004	.0008	.0009
Capitalist society (Kapitalist toplum)	5.13	4.2	5.48	5.04	.071	.001	.014	.0005
Likely result (Olası sonuç)	4.78	2.06	4.22	3.4	.0043	.0007	.0015	.0000
New result (Yeni sonuç)	3.85	3.52	4.3	5.13	-.0004	-.0003	.0000	.0000
Good result (Iyi sonuç)	4.04	6.28	5.14	6	-.000011	-.00007	.0033	.0037
Different result (Farklı sonuç)	3.99	4.27	4.48	5.34	.00049	.00017	.0005	.0002
Nuclear industry (Nükleer sanayi)	6.2	2.02	5.97	2.65	.033	.002	.0006	.0000
New industry (Yeni sanayi)	4.65	4.68	5.43	5.28	.0003	.0003	.0002	.0004
Local industry (Yerel sanayi)	4.87	3.11	5.14	3.07	.0019	.0007	.0008	.0001
Heavy industry (Ağır sanayi)	4.88	4.49	5.27	4.82	.011	.0008	.002	.0005

4.2.2 Identifying adjective-noun collocations' scales of frequency and association.

In addition to comparing frequency and association scales of adjective-noun combinations in English and Turkish, this study identified the scales of RCF and LD-scores in both languages. Tables 4.9 and 4.10 demonstrate the scales of frequency and association counts of unlemmatised adjective-noun combinations in both languages in the high-frequency band. As can be seen in Table 4.9, *working time* with a LD-score of -1.52 is the weakest-associated unlemmatised adjective-noun pair according to the LD-scores in English. This combination was inserted to demonstrate the lowest LD-score of unlemmatised adjective-noun pairs in the high-frequency band of the BNC. The combination *young man* with a LD-score of 7.89 is the strongest-associated unlemmatised adjective-noun collocation according to the LD-scores. This collocation was extracted through node *man* (see Figure 4.12 for RCF and Figure 4.13 for MI-scores). In the high-frequency band, 12 out 20 (60%) unlemmatised adjective-noun combinations have LD-scores of between 4 and 6, and 8 out of 20 (40%) unlemmatised combinations reach LD-scores of between 6 and 8. The collocations which have LD-scores between 4 and 6 obtain RCF scores of between 2.65 and 11.24 PMW. The collocations which have LD-scores between 6 and 8 reach RCF scores of between 7.83 and 46.82 PMW (see Table 4.9).

Table 4.10 demonstrates frequency and association scales of unlemmatised adjective-noun combinations in Turkish in the high-frequency band. As can be seen in Table 4.10, *cyclical time* (çevrimsel zaman) with a LD-score of (-1.15) is the weakest-associated adjective-noun pair according to LD-scores in Turkish. This combination was inserted to Table 4.10 to demonstrate the combination with lowest LD-score in the high-frequency band of the TNC. The combination *following day* (ertesi gün) with a LD-score of 8.19 is the strongest-associated unlemmatised adjective-noun collocation according to the LD-scores. This collocation was extracted through node day (see Figures 4.7 for RCF, and Figure 4.8 for MI-scores). In the high-frequency band, 11 out 20 (55%) adjective-noun combinations have LD-scores of between 4 and 6, and 8 out of 20 (40%) combinations reach LD-scores of between 6 and 8. Furthermore, one collocation *following day* (5%) reaches LD-score of ≥ 8 in Turkish, the strongest-associated collocation in the high-frequency band. It has a RCF score of 59.59 PMW. The collocations which have LD-scores between 4 and 6 obtain RCF scores of between 2.2 and 15.09 PMW. The collocations which have LD-scores between 6 and 8 reach RCF scores of between 5.5 and 28.43 PMW (see Table 4.10).

Table 4.9 Frequency and association scales of unlemmatised high-frequency band collocations in English

LD-points	0	2	4	6	8
RCF scores			2.65-13.82	7.83-46.82	
Percentages	0%	0%	12 out of 20 (60%)	8 out of 20 (40%)	0%
Example collocations	Working time (-1.52)		New day (4.63) Good time (5.18) Outside world (5.81)	Real world (6.1) Long way (6.56) Young man (7.89)	

Table 4.10 Frequency and association scales of unlemmatised high-frequency band collocations in Turkish

LD-points	0	2	4	6	8
RCF scores			2.2-15.09	5.5-28.43	59.59
Percentages	0%	0%	11 out of 20 (55%)	8 out of 20 (40%)	1 out of 20 (5%)
Example collocations	Cyclical time (-1.16)		Real world (4.42) Short time (5.14) New day (5.82)	Whole world (6.78) Old man (7.41) Young man (7.48)	Following day (8.19)

Table 4.11 shows frequency and association scales of English lemmatised adjective-noun combinations in the high-frequency band. As can be seen in Table 4.11, *astonishing time* with a LD-score of -2.73 is the weakest-associated lemmatised adjective-noun pair according to the LD-scores in the high-frequency band. This combination was inserted to Table 4.11 to demonstrate the lowest LD-score of lemmatised adjective-noun pairs in the high-frequency band. The collocation *young man* with a LD-score of 7.77 is the strongest-associated adjective-noun collocation. Similar to the unlemmatised combinations, 12 out 20 (60%) lemmatised combinations have LD-scores of between 4 and 6, and 8 out of 20 (40%) unlemmatised combinations reach LD-scores of between 6 and 8 in the high-frequency band. The collocations which have LD-scores between 4 and 6 reach RCF scores of between 3.08 and 13.78 PMW. The collocations which have LD-scores between 6 and 8 obtain RCF scores of between 7.85 and 47.19 PMW (see Table 4.11).

Table 4.12 shows frequency and association scales of Turkish lemmatised adjective-noun combinations in the high-frequency band. As can be seen in *Table 4.12*, *cyclical time* (çevrimsel zaman) with a LD-score of -1.22 is the weakest-associated lemmatised adjective-noun pair according to the LD-scores in the high-frequency band. This combination was used to demonstrate the lowest LD-score of lemmatised adjective-noun pairs in the high-frequency band. The combination *whole world* (bütün dünya) with a LD-score of 7.48 is the strongest-associated adjective-noun collocation. In the high-frequency band, 8 out of 20 (40%) Turkish lemmatised combinations reach LD-scores of between 4 and 6 in the high-frequency band and 12 out 20 (60%) Turkish lemmatised reach LD-scores of between 6 and 8. The collocations which have LD-scores between 4 and 6 obtain RCF scores of between 7.02 and 26.24 PMW. The collocations which have LD-scores between 6 and 8 reach RCF scores of between 22.51-74.03 PMW (see Table 4.12).

Table 4.11 Frequency and association scales of lemmatised high-frequency band collocations in English

LD-points	0	2	4	6	8
RCF scores			3.08-13.75	7.85-47.19	
Percentages	0%	0%	12 out of 20 (60%)	8 out of 20 (40%)	0%
Example Collocations	Astonishing time (-2.73)		Whole day (4.09) Good man (5.02) Short time (5.35)	Real world (6.08) Other way (6.68) Young man (7.77)	

Table 4.12 Frequency and association scales of lemmatised high-frequency band collocations in Turkish

LD-points	0	2	4	6	8
RCF scores			7.02-26.24	22.51-74.03	
Percentages	0%	0%	8 out of 20 (40%)	12 out of 20 (60%)	0%
Example Collocations	Cyclical time (-1.22)		Other way (4.44) Real world (4.93) Big man (5.98)	Only way (6.08) Following day (7.4) Whole world (7.48)	

Table 4.13 and 4.14 reveal the scales of frequency and association scales of unlemmatised adjective-noun combinations in both languages in the mid-frequency band. As can be seen in Table 4.13, *wrong country* with a LD-score of -0.31 is the weakest-associated English unlemmatised adjective-noun pair according to the LD-scores in the mid-frequency band. This combination was added to Table 4.13 to demonstrate the lowest LD-score of unlemmatised English adjective-noun pairs in the mid-frequency band. The collocation *nuclear industry* with a LD-score of 6.2 is the strongest-associated adjective-noun collocation (see Figure 4.24 for frequency and Figure 4.25 for association counts). In the mid-frequency band, 3 out of 20 (15%) unlemmatised combinations have LD-scores of between 2 and 4, 16 out of 20 (80%) unlemmatised combinations have LD-scores of between 4 and 6. The collocations which have LD-scores between 2 and 4 obtain RCF scores of between 0.88 and 1.49 PMW. The collocations which have LD-scores between 4 and 6 reach RCF scores of 1.08 and 3.53 PMW. The collocation *nuclear industry* with a LD-score of 6.2, the strongest-associated collocation obtains RCF score of 2.88 PMW.

Table 4.14 shows frequency and association counts of Turkish unlemmatised adjective-noun combinations in the mid-frequency band. As can be seen in Table 4.14, *cognate country* (soydas ülke) a LD-score of -0.29 is the weakest-associated Turkish unlemmatised adjective-noun pair according to the LD-scores in the mid-frequency band. This combination was used to demonstrate the lowest LD-score of lemmatised adjective-noun pairs in the high-frequency band. The combination *civil society* (sivil toplum) with a LD-score of 9.94 is the strongest-associated adjective-noun collocation. In the mid-frequency band, 4 out of 20 (20%) collocations have LD-scores of between 2 and 4, 12 out of 20 (60%) collocations have LD-scores of between 4 and 6, and 3 out of 20 (15%) collocations reach LD-scores of between 6 and 8 in Turkish. The collocations which have LD-scores between 2 and 4 obtain RCF scores of between 0.15 and 0.94 PMW. The collocations which have LD-scores between 4 and 6 reach RCF scores of between 0.53 and 3.21 PMW. The collocations which have LD-scores between 6 and 8 reach RCF scores of between 1.95 and 3.07 PMW. The strongest-associated adjective-noun collocation *civil society* (sivil toplum) with a LD-score of 9.94 obtains RCF score of 45.54 in Turkish.

Table 4.13 Frequency and association scales of unlemmatised mid-frequency band collocations in English

LD-points	0	2	4	6	8	
RCF scores			0.88-1.49	1.08-3.53	2.88	
Percentages		0%	3 out of 20 (15%)	16 out of 20 (80%)	1 out of 20 (5%)	0%
Example collocations	Wrong country (-0.31)		New country (3.94) New result (3.85) Different result (3.99)	*Other family (4.36)* *Foreign country (4.9)* Civil society (5.99)	Nuclear industry (6.2)	

Table 4.14 Frequency and association scales of unlemmatised mid-frequency band collocations in Turkish

		0	2	4	6	8
RCF scores			0.15-0.94	0.53-3.21	1.95-3.07	45.54
Percentages		0%	4 out of 20 (20%)	12 out of 20 (60%)	3 out of 20 (15%)	1 out of 20 (5%)
Example Collocations	Cognate country (-0.29)		Nuclear industry (2.02) Local industry (3.11) New result (3.52)	*New family (4.06)* *Other family (4.9)* Whole country (5.35)	Other country (6.06) Foreign country (6.13) Good result (6.28)	Civil society (9.94)

Table 4.15 shows frequency and association scales of English lemmatised adjective-noun combinations in the mid-frequency band. As can be seen in Table 4.15, *wrong country* with a LD-score of -0.88 is the weakest-associated English lemmatised adjective-noun pair according to the LD-scores in the mid-frequency band. This combination was used to demonstrate the lowest LD-score of lemmatised adjective-noun pairs in the high-frequency band. The collocation *other country* with a LD-score of 7.63 is the strongest-associated adjective-noun collocation. In the mid-frequency band, 18 out of 20 (90%) collocations have LD-scores of between 4 and 6, 2 out of 20 (10%) collocations reach LD-scores of between 6 and 8 in English. The collocations which have LD-scores between 4 and 6 reach RCF scores of between 2.18 and 3.84 PMW. The collocations which have LD-scores between 6 and 8 obtain RCF scores of between 3.97 and 27.24 PMW. The strongest-associated adjective-noun collocation *other country* with a LD-score of 7.63 reaches RCF score of 27.24 in English.

Table 4.16 shows frequency and association scales of Turkish lemmatised adjective-noun combinations in the mid-frequency band. As can be seen in Table 4.16, *cognate country* (soydas ülke) with a LD-score of -0.84 is the weakest-associated Turkish lemmatised adjective-noun pair according to the LD-scores in the mid-frequency band. This combination was added to Table 4.16 to demonstrate the lowest LD-score of lemmatised adjective-noun pairs in the mid-frequency band. The collocation *civil society* (sivil toplum) with a LD-score of 8.57 is the strongest-associated adjective-noun collocation. In the mid-frequency band, 3 out of 20 (15%) collocations have LD-scores of between 2 and 4, 12 out of 20 (60%) collocations have LD-scores of 4 and 6, and 4 out of 20 (20%) collocations have LD-scores of 6 and 8 in Turkish. The collocations which have LD-scores between 2 and 4 obtain RCF scores of between 0.33 and 2.19 PMW. The collocations which have LD-scores between 4 and 6 reach RCF scores of between 1.49, and 9.96 PMW. The collocations which have LD-scores between 6 and 8 reach RCF scores of between 5.89 and 50.41 PMW. The strongest-associated adjective-noun collocation *civil society* (sivil toplum) with a LD-score of 7.63 obtains RCF score of 59.13 in Turkish.

. Table 4.15 Frequency and association scales of lemmatised mid-frequency band collocations in English

LD-points	0	2	4	6	8
RCF scores			2.18-3.84	3.97-27.24	
Percentages	0%	0%	18 out of 20 (90%)	2 out of 20 (10%)	0%
Example Collocations	Wrong country (-0.88)		*Likely result (4.22)* New society (4.85) Civil society (5.97)	Modern society (6.16) Other country (7.63)	

Table 4.16 Frequency and association scales of lemmatised mid-frequency band collocations in Turkish

LD-points		0	2	4	6	8
RCF scores			0.33-2.19	1.49-9.96	5.89-50.41	59.13
Percentages		0%	3 out of 20 (%15)	12 out of 20 (60%)	4 out of 20 (20%)	1 out of 20 (5%)
Example collocations	Cognate country (-0.84)		Likely result (3.4) Nuclear industry(2.65) Local industry (3.07)	New family (4.78) Heavy industry(4.82) Other family (5.14)	Good result (6) Foreign country (6.23) Other country (7.17)	Civil society (8.57)

4.3 Discussion

This study extracts four highest frequency adjective collocates of the selected noun nodes in high- and mid-frequency bands within L3-R3 collocation window span. The L3-R3 window span was used to observe the effects of intervening words in the two languages. This choice did not affect the results of the analysis of frequency and association statistics because the correction for window size was used for calculating association counts. Since the focus of the study was on relatively frequently used collocations, low-frequency band node words were not included in this study. The present study considered collocational strength as a continuum of strongly-associated to weakly-associated rather than categorically distinguishing collocations vs non-collocations. The main aim of this inclusive approach to the frequency of occurrence and collocational strength was to observe the frequency-based extracted adjective-noun collocations' scales of frequency and association counts in Turkish and English. The following two sections (4.3.1 and 4.3.2) discuss the scales of frequency and association scores in Turkish and English, and how the collocations were selected for the psycholinguistic experiment presented in the next section (see chapter 6).

4.3.1 The scales of relative collocate frequency and association scores in English and Turkish.

The results show that 11 out of 20 (55%) Turkish unlemmatised adjective-noun combinations, and 15 out of 20 (75%) Turkish lemmatised combinations occur at higher RCF scores than their English equivalents in the high-frequency band. Same as the high-frequency band, 11 out of 20 (55%) Turkish unlemmatised adjective-noun combinations, and 15 out of 20 (75%) Turkish lemmatised combinations occur at higher RCF scores than their English equivalents in the mid-frequency band. This is a surprising finding since unlemmatised forms of the collocations in Turkish present only the base forms whereas the unlemmatised forms in English potentially subsumes the equivalents of the base, as well as the case, instrumental and person inflected forms of the combinations in Turkish. It could therefore be expected to see a majority of English unlemmatised adjective-noun collocations to occur at higher RCF scores than their Turkish equivalents. However, the results show that unlemmatised adjective-noun combinations in both languages tend to occur at similar RCF scores. For example, high-frequency band English unlemmatised adjective-noun combinations reach a mean RCF score of 12.64 (SD=10.97) with the maximum of 46.82 (*long time*), and the minimum of 2.65 (*whole day*). Turkish unlemmatised adjective-noun combinations reach a mean RCF score of 14.97 (SD=13.38) with the maximum of 59.59 (Following day), and the minimum of 2.2 (Real world). The results also indicate that the differences between word pairs' unlemmatised and lemmatised RCF scores are predominantly larger in Turkish than in English since in both frequency bands 15 out of 20 (75%) Turkish lemmatised combinations occur at higher RCF scores than their English equivalents. This is an expected finding since lemmatised collocations present the base and plural forms of the collocations only whereas in Turkish the lemmatised forms involve base, as well as 5 different case-marked, and plural forms (for examples see Table 4.5 Table 4.6).

In addition to the RCF scores, it is important to explore collocational strength in English and Turkish using the MI and LD-scores to have a broader picture of adjective-noun collocations in the two languages. The results show that 8 out of 20 (40%) Turkish unlemmatised adjective-noun pairs, and 10 out of 20 (50%) Turkish lemmatised adjective-noun pairs reach higher MI-scores than their English equivalents in the high-frequency band. In the mid-frequency band, 10 out of 20 (50%) Turkish unlemmatised adjective-noun combinations reach higher MI-scores, and 12 out of 20 Turkish lemmatised adjective-noun combinations have higher MI-scores than their English equivalents. I also looked at the mean MI-scores of English and Turkish combinations in the both frequency bands. In the high-frequency band, the unlemmatised English combinations have a mean MI-score of 2.04 (SD=1.38), and when they are lemmatised, the mean MI-score decreased to 1.91 (SD=1.36). The unlemmatised Turkish combinations obtain mean MI-score of 1.91 (SD=1.77), and when they are lemmatised, the mean MI-score decreased to 1.83 (SD=1.6). In the mid-frequency band, the unlemmatised English combinations have a mean MI-score of 1.68 (SD=1.96), and the lemmatised English combinations have a mean MI-score of 1.51 (SD=1.65). The unlemmatised Turkish combinations obtain a mean MI-score of 1.49 (SD=1.91), and lemmatised Turkish combinations reach a mean MI-score of 1.57 (SD=1.73). Based on these findings, it is possible to say that a majority of the lemmatised Turkish collocations appears to have higher MI-scores than their English equivalents in the mid-frequency band. Nevertheless, it should be noted that the collocational relationship is a complex one and no single measure of association could explain the full complexity of this relationship (Brezina et al. 2015). Especially considering the limitations of MI-scores (see Gablasova et al. 2017 and section 3.1.1 for a discussion on selecting association measures), it is important to look at the scales of LD-scores languages.

As previously mentioned (see section 4.2.1), the mean LD-scores show that English lemmatised combinations obtain a lower mean LD-score of 5.72 (SD=1.05) than the unlemmatised combinations 5.83 (SD=1), whereas Turkish lemmatised combinations reach a slightly higher mean LD-score of 6.15 (SD=0.87) than the unlemmatised combinations 6.07 (SD=1.06) in the high-frequency band. To observe the scales of unlemmatised and lemmatised combinations' LD-scores in both languages in the high-frequency band, I closely examined Table 4.*7*, andTable 4.*8*. Only one unlemmatised Turkish collocation reaches an LD-score of ≥8 *ertesi gün* (following day) (LD=8.19), and no unlemmatised English collocation obtains an

LD-score of ≥8 in the high-frequency band. Moreover, in English 12 out of 20 (60%) unlemmatised combinations have LD-scores between 4 and 6. These twelve collocations have a mean RCF score of 6.48 (SD=3.26), with the maximum of 13.82 (*short time*), and the minimum of 2.65 (*whole day*) in English. In Turkish 11 out of 20 (55%) unlemmatised combinations have LD-scores between 4 and 6. These eleven collocations have a mean RCF score of 7.59 (SD=4.35), with the maximum of 15.09 (*good time*), and the minimum of 2.2 (*real world*). Furthermore, 8 out of 20 (40%) unlemmatised combinations have LD-scores between 6 and 8 in both languages. The eight collocations in English have a mean RCF score of 21.87 (SD=12.14) with the maximum of 46.82 (*long time*), and the minimum of 7.83 (*real world*). The eight collocations in Turkish with LD-scores of between 6 and 8 have a mean RCF score of 19.68 (SD=7.79) with the maximum of 28.43 (*previous day*) and the minimum of 5.5 (*good man*). Looking at the scales of RCF and LD-scores for high-frequency band combinations, it is possible to say that distribution of RCF and LD scores for unlemmatised combinations in both languages are quite similar.

No lemmatised collocation in English and Turkish reaches an LD-score of ≥8 in the high-frequency band. As can be seen in Table 4.7, 12 out of 20 (60%) lemmatised combinations have LD-scores between 4 and 6 in English. These twelve collocations obtain a mean RCF score of 7.48 (SD= 3.65) with the maximum of 14.09 (*short time*), and the minimum of 3.08 (*whole day*) in English. In Turkish, 8 out of 20 (40%) lemmatised combinations have LD-scores between 4 and 6. They reach a mean RCF score of 14.26 (SD=6.24) with the maximum of 26.24 (*good time*), and the minimum of 7.02 (*different way*). In English, 8 out of 20 (40%) whereas in Turkish 12 out of 20 (60%) lemmatised combinations have LD-scores between 6 and 8. The English collocations with LD-scores of between 6 and 8 obtain a mean RCF score of 26.22 (SD=12.78) with the maximum of 47.19 (*long time*), and the minimum of 7.85 (*real *world*). The Turkish collocations with LD-scores of between 6 and 8 obtain a mean RCF score of 40.08 (SD=16.11), with the maximum of 74.03 (*following day*), and the minimum of 22.51 (*only way*). Based on the distribution of LD and RCF scores in both languages, it is clear that lemmatised combinations in Turkish are considerably more frequent and strongly associated than the unlemmatised combinations in the high-frequency band.

Looking at the scales of RCF, and association counts in each frequency bands is a very useful approach to see the general picture of how lemmatisation affects collocational strength in typologically different languages English and Turkish. Nevertheless, it is useful to investigate the extent to which this general trend followed by all of the collocations in the high-frequency band. On the one hand, the Turkish collocation, *kısa zaman* (short time) follows this general trend since it obtains higher MI-score of 2.8 and LD-score of 6.56 than its unlemmatised form (1.41 and 5.14 respectively). As shown in Figure 4.3, the locative inflected form *kısa zamanda* is a considerably more frequent and strongly associated than other case marked forms of the collocation. To understand the reason why the lemmatised collocation *kısa zaman* (short time) is a considerably more frequent and strongly associated collocation than its unlemmatised form in Turkish, I looked at RCF and association scores of the locative inflected form *kısa zamanda* (*short time-in*), which obtains RCF score of 36.11 and LD-score of 8.97. The concordance lines of the locative inflected form *kısa zamanda* display the frequent use of a formulaic expressions *en kısa zamanda* (most short time-in) with RCF score of 1.71. This formulaic expression is the meaning-equivalent of *as soon as possible* in English. On the other hand, another Turkish collocation *yeni dünya* (new world) is an example for collocations whose unlemmatised form reaches higher association scores than its lemmatised form. It obtains an RCF score of 23.04 and its lemmatised form reaches an RCF score of 39.72 in Turkish. Surprisingly the lemmatised collocation *yeni dünya* obtains lower MI-score of 1.41 and LD-score of 6.85 than its unlemmatised form (2.05 and 7.19 respectively). That is to say, despite the increasing relative collocation frequency score, the lemmatised form of *yeni dünya* does not appear to be as strongly associated as its unlemmatised form. Furthermore, as can be seen in Figure 4.6, the case and plural inflected forms seem to be considerably weaker associated than the base form of this collocation.

The collocation *ertesi gün* (following day) is another example for collocations whose unlemmatised form reaches higher association scores than its lemmatised form. The Turkish unlemmatised collocation *ertesi gün* reaches a considerably lower RCF score of 59.59 than its lemmatised form (74.03). Surprisingly unlemmatised collocation *ertesi gün* reaches a considerably higher MI-score of 6.52 and LD-score of 8.19 than its lemmatised form (5.71 and 7.4 respectively). The RCF and LD score of the case inflected and plural forms of the collocation *ertesi gün* in Turkish show that the inflected forms are both less frequent and weaker associated than the base form. Similar to the collocation *yeni dünya* despite the lemmatised form of this collocation's high-frequency as measured by RCF scores, it is weaker associated than its unlemmatised form. Furthermore, the concordance lines are a very helpful resource to explore why the collocation *ertesi gün* reaches considerably higher LD-scores than its English equivalent *following day*. The concordance lines of the adjectives *following* in English and *ertesi* in Turkish reveal that the adjective *ertesi* in Turkish is predominantly used to modify nouns such as *day*, *morning*, *night*, *evening*, and *week,* whereas *following* is used to modify wide variety of nouns including *day*, *year*, *government*, *decision* and *accident* in English. That is to say, *following* is used in a more general context and with wider variety of nouns than its Turkish equivalent *ertesi*. As a result of this, the adjective *following* reaches considerably higher relative frequency 269.05 than its Turkish equivalent (99.09). The adjective *ertesi* in Turkish contributes to the exclusiveness of the collocation *ertesi gün* considerably more than the adjective *following* in English.

As can be seen in Table 4.8, only one lemmatised Turkish collocation reaches LD-score of ≥8 in the mid-frequency band. The collocation *sivil toplum* (civil society) obtains a LD-score of 8.57, and RCF of 59.13 in Turkish. No lemmatised English collocation reaches an LD-score of ≥8 in the mid-frequency band. Furthermore, no lemmatised English collocation obtains LD-scores of between 2 and 4, but 3 out of 20 (15%) lemmatised Turkish collocations have LD-scores between 2 and 4. These 3 collocations have a mean RCF score of 0.99 (SD=1.03) with the high score of 2.19 (*new result*), and the low score of 0.33 (*nuclear industry*). In English 18 out of 20 (90%) lemmatised collocations have LD-scores between 4 and 6. They reach a mean RCF score of 2.72 (SD=0.69) with the high score of 3.84 (*other family*), and the low score of 1.59 (*likely result*). In Turkish, 12 out of 20 (60%) lemmatised Turkish collocations have LD-scores between 4 and 6. They obtain a mean RCF score of 7.96 (SD=4.31) with the high score of 18.35 (*new country*), and the low score of 1.67 (*heavy industry*). Moreover, 2 out of 20 (10%) lemmatised collocations have LD-scores between 6 and 8 in English. These collocations *other country* and *modern society* have a mean RCF score of 15.6 (SD=16.45). In Turkish 4 out of 20 (20%) lemmatised collocations obtain LD-scores between 6 and 8. They have a mean RCF score of 28.75 (SD=15.02) with the high score of 50.41 (*other country*), and the low score of 15.66 (*good result*). These findings suggest that Turkish lemmatised collocations in the mid-frequency band appear to be predominantly more frequent and strongly associated than their Turkish equivalents.

Alongside the general trend of positive effect of lemmatisation on the collocational strength of Turkish adjective-noun collocations, it is important to look at the examples of the unlemmatised collocations which obtain lower MI and LD scores than their lemmatised forms in the mid-frequency band. The collocation *sivil toplum* (civil society) is an example for some unlemmatised collocations' having higher association scores than their lemmatised forms in Turkish. It obtains RCF score of 45.54 and LD-score of 9.94. When it is lemmatised the collocation *sivil toplum* reaches an RCF score 59.13. However, both LD and MI scores are lower 6.57 and 8.57 than the unlemmatised form (8.03 and 9.94 respectively). Exploring the association counts of the inflected forms of this collocation in the two languages revealed that the case inflected and plural forms in Turkish appear to be weaker associated word pairs than the base form of this collocation in Turkish. Therefore, the lemmatised collocation *sivil toplum* obtains lower association scores than their unlemmatised forms in Turkish. It is also important to point out that there is a large difference between RCF and association scores of the collocation *civil society* (sivil toplum) in English and Turkish. Therefore, I consulted the concordance lines in English and Turkish to understand the nature of the large differences of RCF and association scores in Turkish and English. The concordance lines suggest that the three-word constructions with similar meanings, *sivil toplum kuruluşu* (civil society foundation) and *sivil toplum örgütü* (civil society organisations) are frequently used in Turkish. They are the direct equivalents of the construction Non-governmental organisation in English. That is to say, the collocation *sivil toplum* is frequently used as constituents of these three-word, high-frequency constructions in Turkish. This is one of the reasons for this large frequency and association differences of this collocation in Turkish and English.

This study sets out to explore frequency and association counts of adjective-noun collocations in English and Turkish, and whether agglutinating structure of Turkish affects the collocability of adjectives and nouns. High- and mid-frequency band unlemmatised collocations have similar mean relative collocation frequency scores (RCF) and association counts as measured

To observe the effect of agglutinating structure of Turkish on the collocability of adjectives and nouns, the lemmatised forms of the collocations in the both languages were compared. The results suggested that the vast majority of the lemmatised Turkish adjective-noun combinations occur at a higher-frequency than their English equivalents. Looking at the association scores, MI-scores show that only mid-frequency band Turkish lemmatised collocations have higher mean MI-scores than their unlemmatised forms. High-frequency band lemmatised collocations in both languages obtain lower mean MI-scores than their unlemmatised forms. This finding can be interpreted as agglutinating structure of Turkish affects the high-frequency and mid-frequency bands combinations differently. However, it is important to bear in mind that MI-score has been found to favour low-frequency combinations (Ebeling & Hasselgård, 2015; Gablasova et al. 2017), therefore it is not the most ideal measure of association to observe the collocational strength of high-frequency lemmatised collocations. LD-scores show that agglutinating structure of Turkish appears to increase adjective-noun combinations' collocational strength in the both frequency bands. Nevertheless, it should be noted that this finding cannot be generalised to all of the word pairs, since some unlemmatised collocations reach higher MI and LD scores than their lemmatised forms. That is to say, agglutinating structure of Turkish does not increase the collocational strength of all Turkish collocations (e.g. *new world* and *following day*). The reason is that the inflected form of some of these collocations are considerably weaker-associated than the base forms of these collocations.

It should be noted that these findings related to the crosslinguistic differences of adjective-noun collocations' frequency in English and Turkish are in line with the previous corpus findings from contrastive studies. For example, Cortes (2008) found that while noun-noun pre-modifications are more frequent in English (e.g. *immigration history*), than they are in Spanish. In Spanish, they need to be expressed by a post modifying prepositional phrase (e.g. *la historia* *de la immigración*). Therefore, there are more four-word lexical bundles in Spanish corpus than English corpus. Furthermore, Granger (2014) also found marked differences between the type and frequency of lexical bundles found in parliamentary and newspaper editorial genres in English and Turkish. To conclude, this section (4.3.1) focused on the frequency-based extracted adjective-noun collocations' scales of RCF and association scores in high-mid-frequency bands, in English and Turkish. The next section (4.3.2) discusses the link between the corpus-based analysis focusing on the scales of RCF and association and psycholinguistic experiment presented in chapter-3.

4.3.2 Linking corpus-based analysis to psycholinguistic reality.

The inference from corpus to mind about processing of collocations is mainly based on two assumptions. First, the frequency of the particular features experienced in the language system have an influence on the cognitive representation of those features (Ellis, 2008, 2012, Ellis et al. 2015). Second, words in a collocational relationship can be said to predict one another in that the presence of one word makes the presence of other word more likely (Hoey, 2005: p. 6-7). These assumptions seem to be in line with one of the main tenets of the usage-based models of language, which suggests that language system is shaped by the frequency of occurrence across all linguistic levels (Christiansen & Chater, 2016, Ellis, 2008, 2012, Kemmer & Barlow, 2000). The advantage of using corpora is that it can provide direct information not only on frequencies of the constituents of the collocations but also their co-occurrence frequency, and strength of associations related to dispersion, exclusivity and directionality (Brezina et al. 2015, Ellis et al. 2015, Evert, 2008, Gablasova et al. 2017b, Gries, 2013). In this regard, corpora as large databases documenting the regularities in the use of collocations (Gablasova et al. 2017b), have a potential to contribute to the psycholinguistic studies for specifying the factors influencing the cognitive processing and representation. This section discusses the frequency of occurrence and collocational strengths as corpus related variables affecting the mental processing and representation of adjective-noun collocations by L1 and L2 speakers.

The psycholinguistic experiments (reported in Chapter 6) explore the processing of high-frequency and low-frequency adjective-noun collocations in English and Turkish by L1 speakers of these languages. In addition, it reports on the processing of adjective-noun collocations in English by L1 English and L1 Turkish-English L2 advanced speakers. More specifically, they firstly look at the effects of the factors single word frequency of adjectives and nouns and collocational frequency counts on the mental processing of adjective-noun collocations in English and Turkish for L1 speakers of these languages. Considering the corpus-based findings provided by the current chapter that agglutinating structure of Turkish tend to increase adjective-noun pairs' collocational strength in high-mid-frequency bands (see section 4.3.1 for summary of the corpus findings), the psycholinguistic experiments therefore explore the extent to which L1 speakers of these languages show sensitivity to single word and collocational frequency counts differently in English and Turkish. Secondly, another experiment examines whether L1 Turkish-English L2 advanced speakers show sensitivity to single word ad collocational frequency counts in their L2.

Looking at the distributions of RCF and association statistics of base and inflected forms of the adjective-noun collocations in Turkish, it is possible to say that the frequency and association counts' distributions of the unlemmatised Turkish collocations are quite similar to the English ones. It is possible to say that the usage-based approaches to language acquisition (Bybee, 1998; Christiansen & Chater, 2016a; Ellis, 2002; Goldberg, 2006; Kemmer & Barlow, 2000; Tomasello, 2003) would be consistent with the both English and Turkish adjective-noun collocations, since speakers of languages are likely to be sensitive to the frequency information at multiple grain sizes including the frequencies of the single words within the collocations and the whole collocations. Given the fact that unlemmatised Turkish and English collocations have similar RCF and collocational strength counts, they are both likely to be sensitive to the collocational frequency counts of the collocations. Furthermore, it is important to bear in mind that Turkish nouns can be inflected with various markers such as case, so that Turkish speakers are also likely to attend to single word frequency counts of the nouns. As Durrant (2013) notes that high-frequency morpheme bundles are not neutral with regard to the lexis. Therefore, speakers of Turkish are also predicted to be sensitive to the frequencies of the inflected forms of the nouns alongside their base forms.

One important point to be discussed from a corpus linguistics perspective is that how the collocations in English and Turkish should be extracted from the BNC and TNC for the experiments. Many psycholinguistic studies used MI-score of 3 as a threshold to extract significant (Wolter & Yamashita, 2015), tight (González Fernández & Schmitt, 2015), coherent (Ellis et al. 2008), and appropriate (Siyanova & Schmitt, 2008) word combinations. The MI-score is an effect size measure, and it does not test how much evidence the corpus provides for a significance of word combinations co-occurrences. Therefore, the threshold used by many psycholinguistic experimental works (e.g. Wolter & Yamashita, 2015, Vilkaite, 2016), the MI-score of 3 does not indicate any significance related to the word combinations' co-occurrences. Furthermore, the other descriptions used by psycholinguistic studies (e.g. coherence, appropriateness, tightness) do not seem to be clear and transparent for what aspect of the collocations they highlight. Since the current study specifically aims to select high- and low-frequency collocations in both languages, using MI measure would not be the ideal way of extracting frequent collocations. The main reason is that the MI measure favours low-frequency combinations, particularly the word combinations with technical meaning in a large corpora (Evert, 2008; Gablasova et al. 2017). For example, I checked one of the node words *industry* to see the RCF scores of adjective-noun pairs with high MI-scores. The word pair *automative industry* (MI=7.5), obtains a RCF score of 0.29, and *petrochemical industry* (MI=7.32), has a RCF score of 0.19.

To be able extract high and low-frequency collocations in a systematic way, the psycholinguistic experiments (see Chapter 6) used LD measure. The main reason for preferring LD-score is that it provides a chance to highlight exclusive collocations without the low-frequency bias, and with a clearly delimited scale (Gablasova et al. 2017). To determine the threshold values for high and low-frequency adjective-noun pairs, I relied on the scales of adjective-noun pairs presented in Table 4.5,Table 4.*6*, Table 4.*7*, and Table 4.*8*. To extract the low-frequency collocations, I selected the adjective-noun pairs with LD scores of ≤ 4 and ≥ 2 from the BNC and TNC, for English and Turkish respectively. I extracted a total of 30 low-frequency collocations for English (e.g. *lovely house*, *warm place*, *elderly mother*). They had a mean RCF score of 0.55 (SD=0.27) and LD score of 3.24 (SD=0.27). To extract these collocations, I used the nouns between the relative frequency of 512.45 (*world*), and 83.29 (*glass*). To extract the high-frequency collocations, I selected the adjective-noun pairs with LD scores of ≥7. I extracted a total of 30 high-frequency collocations for English (e.g. *white paper*, *front door*, *dark hair*, *medical treatment*) (see appendix E and F for a full list of items). They obtain a mean RCF score of 15.72 (SD=16.97), and LD score of 7.8 (SD=0.83). Using the same threshold values, I extracted 26 low-frequency and high-frequency collocations in Turkish (see section 6.1.2 for a more detailed explanation for the experiment items).

4.3.3 Limitations.

One important limitation of this corpus study is that it only included the congruent adjective-noun combinations that have translation equivalents in the both languages. The reason for looking only at the congruent collocations was to compare the frequency and collocational strength of the word combinations that have very similar meanings in English and Turkish. Therefore, this study was not in a position to compare the word combinations that have totally different meanings. However, a drawback of this methodological choice is that it limits the scope of the corpus analysis with congruent collocations only and excludes the incongruent word combinations that does not have translation equivalents in the two languages from the analyses. It would be useful for a future corpus study to include the incongruent collocations in the frequency and association scales to obverse the extent to which the scales differ from the ones provided in this study. Another limitation of the current study was that the BNC and the TNC had different approaches to tokenisation. The BNC takes punctuation marks as characters, but the TNC does not. In other words, the BNC and the TNC are not fully comparable in terms of the tokenisation. Therefore, this study took the size of the BNC that excluded the punctuation marks (98,560,118 words). It is important to note that some collocations are considerably more frequent in Turkish than their equivalents in English (e.g. *civil society*). This could also be because of how the TNC chose its texts for certain genres.

4.4 Conclusion

To conclude Part 2 of the thesis firstly provided a detailed literature review on properties of collocations (see section 3.1), contrastive studies on MWS (see section 3.2) and MWS in agglutinating languages (see section 3.3). Secondly, it reported on a contrastive corpus-based study on the adjective-noun collocations in English and Turkish (see chapter 4). The aim of this study was to compare adjective-noun collocations for frequency and collocational strength in the two languages. In this study, working with English and Turkish provided a chance to explore in what ways language typology, particularly agglutinating structure of Turkish language affects the collocability of words. The results showed that agglutinating structure of Turkish appears to increase adjective-noun combinations' collocational strength in the both frequency bands. Nevertheless, it should be noted that this finding cannot be generalised to all of the word pairs, since some unlemmatised collocations reach higher MI and LD scores than their lemmatised forms. The Part 3 (see Chapter 6) reports on two psycholinguistic experiments, it firstly explores the factors affecting the processing of high and low-frequency collocations in English and Turkish. It secondly investigates the processing of high and low-frequency adjective-noun collocations' processing by L1 English and L1 Turkish-English L2 speakers.

Part 3: Processing Multi-word Sequences in Typologically Different Languages

Chapter 1 of this thesis provided a detailed discussion of the main features and the role of MWS, in language processing and representation. To recap briefly, the study of MWS, also known as formulaicity, is closely associated with usage-based models of language (Ellis, 2002, 2008, Christiansen & Chater, 2016, Kemmer & Barlow, 2000). According to these models, a speaker's language system is closely shaped by their life-time experiences of language. Specifically, each experience of language processing (both comprehension, and production) provides an acquisition and processing opportunity. In other words, a fundamental aspect of language acquisition is learning how to process utterances (Christiansen & Chater, 2016; McCauley & Christiansen, 2015). Language processing takes place here-and-now. That is, at a rate of normal speech, L1 speakers of English produce an average of 150 words per minute. Given the very fleeting nature of language input, language users need to perform some form of chunking operations to process the information in real time. For the efficiency of processing, we therefore need to use a wide range of MWS. They include a broad range of constructions that fulfill a number of communicative functions such as collocations, lexical bundles and binomials (Wray, 2002, 2008). Two features are common to all types of constructions that are considered under the heading of MWS. The first is that they are recurrent. The second is that they are processed faster than novel control phrases (Carrol & Conklin, 2019; Vilkaite, 2016). In what follows chapter 5 reviews the studies focussing on the processing of various MWS in different languages and by L1 and L2 speakers.

Chapter 5: Literature Review of Psycholinguistic Experiments

Language users have been shown to be sensitive to the frequency of linguistic units along the continuum from the smallest units (e.g. phonemes) to the largest units (e.g. lexical bundles) (see Ellis, 2002; Christiansen & Chater 2016a for a review). This is compatible with the usage-based approaches to language acquisition which posit that linguistic knowledge is closely related to the lifetime experiences of language (e.g. Kemmer & Barlow, 2000; Tomasello, 2003). More specifically, these approaches view that frequency plays an important role in the processing of linguistic units of various sizes. The previous empirical studies focussed on the frequency effects on the processing of single words and found clear effect of distributional properties (e.g. Rayner & Duffy, 1986). More recently, psycholinguistic experiments provided evidence that language users are sensitive to the frequency effects on linguistics units larger than words. This chapter provides a detailed review of studies focussing on L1 and L2 processing of MWS (see sections 5.1 for L1 processing and 5.2 for L2 processing of MWS). Although this thesis is primarily interested in the processing of collocations, this chapter also reviews studies focusing on other type of MWS (e.g. lexical bundles) alongside collocations to adopt a broader perspective of processing MWS.

5.1 Factors affecting processing of MWS in L1s

The frequency-based corpus linguistic perspective (see, Brezina et al. 2015; Evert, 2008), views collocations as lexical items that co-occur with greater than random probability in corpora (Hoey, 2005, p. 3-5). On this basis, Durrant and Doherty (2010) investigated whether English high-frequency collocations also have a psychological reality. They conducted two experiments using lexical decision tasks (LDT). In the first experiment, they employed a design in which the prime word was presented for 600 milliseconds (ms). Four different item types were used: low-frequency combinations (e.g. *direct danger*); moderate collocations (e.g. *greater concern*); frequent collocations (e.g. *foreign debt*); and associated frequent collocations (e.g. *estate agent*). The items were extracted from the BNC, and they were matched with 48 control items, which had no attestation in the BNC. The participants were thirty-two adult L1 of English. Durrant and Doherty (2010) found no statistically significant difference between collocation and non-collocation conditions in the low-frequency and moderate collocations. However, they found a priming effect in the processing of frequent and associated frequent conditions in comparison to non-collocations. In the second experiment, the same items were tested in a LDT in which the prime word was presented for a considerably shorter duration (60 ms), to preclude any conscious strategies participants possibly used in the first experiment. They found a priming effect in the processing of associated frequent conditions in comparison to non-collocations, however they found no facilitation effect in the low, moderate and frequent conditions. They concluded that frequent but non-associated collocations exhibited a priming effect only in an experimental set-up which allowed the inclusion of strategic processes.

A few studies have looked at the MWS beyond the bigram level. Arnon and Snider (2010) examined if language users are sensitive to the frequencies of compositional four-word phrases controlling for the frequency of the individual words forming the four-word phrases. They conducted two experiments using phrasal decision tasks (PDT), that is participants saw four-word phrases and they were asked to decide if they were possible in English. The participants were adult L1 speakers of English. The first experiment involved 28 items (16 high-frequency and 12 low-frequency). The high- and low-frequency phrases were matched for the frequency of the individual words, but they differed in the phrasal frequency (e.g. *don't have to worry* vs *don't have to wait*). They found that participants' responses were significantly faster for more frequent items in both high-frequency and low-frequency phrases. In the second experiment, they looked at the effect of a group of mid-frequency items on the phrasal frequency. For this group, they constructed 17 items, which had lower phrasal frequency than the high-frequency group and higher phrasal frequency than the low-frequency group. The results of the second experiment showed that participants were faster to respond to mid-frequency phrases than the low-frequency phrases. As additional analyses, they performed a meta-analysis of the items used in experiments in 1 and 2, in order to examine if actual frequency predicted participants' reaction times (RTs), and to see if phrasal frequency was a better measure than a binary frequency measure. The results indicated that the phrasal frequency was a significant predictor of RTs, and when phrasal frequency was taken into account as a continuous predictor, the binary measure was no longer significant. Arnon and Snider (2010) concluded that the results were in accordance with the usage-based models of language where every additional occurrence of sequence strengthens its activation. One important limitation to this study was that high-frequency and low-frequency phrases had quite different semantic structures so that they are likely to create very different expectations about the upcoming information. Furthermore, some of the phrases used in the experiments can be used in isolation, which are fully meaningful (e.g. *how do you feel*), but some others cannot be used in isolation (e.g. *to have a lot*), because they are less meaningful sentence fragments. Jolsvai, McCauley, Christiansen (2013) provided empirical evidence that meaningfulness affects processing speed of MWS. In this regard, Arnon and Snider's (2010) results should be viewed with caution.

There is surprisingly little research focusing on the factors other than the frequency of MWS. Jolsvai, McCauley, Christiansen (2013) investigated the processing of MWS that vary in the degree to which language users find them meaningful as a unit, while controlling for single word and phrasal frequency counts. Three types of items were extracted for the experiment from two different corpora; 3-word idiomatic expressions (e.g. *over the hill*), highly meaningful compositional phrases (e.g. *had a dream*), and less meaningful sentence fragments (e.g. *by the postal*). Both idiomatic expressions, and compositional phrases were rated as being equally meaningful in an initial norming study. The sentence fragments were rated as considerably less meaningful than both the idiomatic and compositional phrases. In a second norming study, a different set of participants rated these three types of items as equally plausible as part of a sentence in English. Forty adult, L1 speakers of American English were recruited for this study. They used the PDT which was previously used by Arnon and Snider (2010). The results indicated that the RTs were faster for the meaningful items (idiomatic and compositional phrases). It took more time for participants to respond to fragments than idiomatic and compositional phrases. However, participants' RTs for compositional phrases were not significantly slower for compositional phrases than for the idiomatic phrases. Phrase frequency also significantly predicted the participants' RTs. Jolsvai, et al. (2013) concluded that participants were sensitive to the meaningfulness of the MWS since they responded to the compositional, and idiomatic phrases, which were rated meaningful, faster than the frequency matched sentence fragments, which were rated as less meaningful. The phrasal frequency counts also significantly predicted the participants RTs, but to a lesser degree than the meaningfulness of the MWS. This study also has an important methodological implication for the future studies which aim to investigate the effect of frequency in the processing of MWS. That is, the meaningfulness of the items should be taken into consideration.

To the best of my knowledge, only one study looked at collocational processing in an agglutinating language. Cangir, Büyükkantarcioglu, and Durrant (2017) focused on the research questions; whether there is any evidence of collocational priming in Turkish collocations, and whether frequency plays any role in the processing of Turkish collocations.

They used a LDT in which the prime word was presented for 100 ms. Using an MI score of 3, and a t-score of 2 as cur-off points, they extracted two different types of items, 60 collocational (e.g. *soğuk savaş* – "cold war") and 60 non-collocational (e.g. *uzak savaş – "far war"*). The items included 30 adjective-noun, and 30 noun-verb (e.g. *hata yapmak – "make a mistake"*) collocations. They used the same nouns, but different adjectives and verbs in collocational and non-collocational conditions. They did not include any inflected forms of either types of items. Cangir et al. (2017) found that there is evidence for collocational priming in Turkish. The items in collocational conditions were responded to faster than the items in non-collocational conditions. The verb-noun collocations were responded to faster than the adjective-noun collocations. Furthermore, they found that the part of speech (adjective-noun vs noun-verb collocations), and target word frequency significantly affected the participants' RTs. This study had a few serious shortcomings. Firstly, single-word and collocational frequency counts do not seem to be systematically controlled. More specifically, the frequency of the prime and target words between collocational and non-collocational conditions, and within adjective-noun and verb-noun collocations do not seem be closely matched. Secondly, although the authors claimed that the participants did not use any conscious strategies in their responses, their experimental set-up allowed the inclusion of strategic processing. The presentation of the prime for 100 ms is above the cut-off duration for the masked-priming paradigm (Jiang, 2012, p. 100). In this regard, the authors did not minimise the possibility of participants' relying on strategic processing. It is therefore important to view the results of this study with caution.

A few studies recorded participants' eye-movements to examine the processing of MWS. McDonald and Shillcock (2003) investigated whether readers are sensitive to the collocational strength, also known as the transitional probabilities of verb-noun collocations (the first experiment). They extracted 48 verbs, and each verb was paired with a highly predictable and a less predictable noun object from the BNC. The transitional probabilities of the items were calculated using the formula *P*(noun|verb)=[frequency(verb,noun)/frequency(verb)]. The mean values were .01011 for the high-predictability, and .00038 for the low-predictability pairs. The length and corpus frequency of the nouns were matched in two conditions. A group of L1 speakers of English rated the sentences in two conditions as equally plausible. Sentences were constructed for high- and low-predictability sentences with an identical neutral context (e.g., high-predictability-*One way to avoid confusion is to make the changes during the vacation*, low-probability-*One way to avoid discovery is to make the changes during the vacation*). Adult L1 English participants took part in the eye-movement experiment. McDonald and Shillcock (2003) found that initial-fixation duration (an early measure of eye-movement) was significantly shorter for verb-noun collocations with a high-predictability than low-predictability pairs. Furthermore, they argued that participants were sensitive to the frequency distribution of verb-noun collocations. The eye-movement data provided empirical evidence for readers' exploiting available statistical information about word-to-word contingencies to predict the upcoming words in English. They argued that the effects of transitional probabilities are independent from contextual predictability.

Replicating and expanding McDonald and Shillcock's (2003) article, Frisson, Rayner, Pickering (2005) examined whether the effects of transitional probabilities of the collocations is totally independent from the effects of contextual predictability. Using the same verb-noun pairs that McDonald and Shillcock (2003) previously used, Frisson et al. (2005) firstly attempted to replicate McDonald and Shillcock's (2003) findings. The verb-noun pairs were preceded by either a constraining context or a neutral context. Frison et al. (2005) found effect of contextual predictability for gaze duration analysis. In line with McDonald and Shillcock (2003), Frison et al. (2005) also found that nouns that follow a verb with transitional probability were read faster, according to the gaze duration measure. Secondly, Frisson et al. (2005) attempted to determine whether the effects of transitional probabilities are still present when the neutral and constraining contexts are better controlled than the first experiment. They used a modified version of the cloze task to assess the predictability of the contexts used and they were able to control the imbalance between high and low-transitional probability items. For their second eye-movement experiment, they prepared a total 56 verb-noun combinations. Some of these verbs were matched with different nouns than the first experiment. They found a significant effect of constraining context leading to shorter reading times, according to the gaze duration analysis. Frison et al. (2005) concluded that transitional probabilities have no significant effect on the collocational processing, if the contextual predictability is well controlled.

Examining the different aspects of collocational processing, Vilkaite (2016) also looked at participants' eye-movements to test if non-adjacent collocations also show processing facilitation as the adjacent ones, controlling for the contextual predictability. L1 speakers of English participated in the experiment. Four groups of items (verb-noun pairs) were prepared for the study: adjacent collocations (e.g. *provide information*); non-adjacent collocations (e.g. *provide some of the information*); adjacent controls (e.g. *compare information*); non-adjacent control (e.g. *compare some of the information*). The verb-noun collocations for the study were extracted from the BNC, using the MI score of 3. The same nouns were retained in all four conditions, and the same intervening words were inserted in both non-adjacent collocations, and non-adjacent control conditions. The individual frequency, and length of the word were closely matched across the four conditions. To control for predictability, the sentences included only a very neutral context. In this study, both the single words; verbs and nouns as single words, and also whole phrases (verb-noun pairs) were defined as areas of interests. Vilkaite

(2016) found that adjacent collocations were read significantly faster than adjacent controls, whereas the effect was not significant when the non-adjacent collocations were compared with non-adjacent controls, according to the gaze duration measure. Furthermore, according to the total reading time measure, a significant facilitation effect was found for the adjacent collocations when compared to adjacent controls. However, no facilitation effect was found when non-adjacent collocations were compared with non-adjacent controls. When looking at the whole phrase reading measures: first pass reading time; fixation count; total reading time, the phrases containing the non-adjacent collocations were consistently read faster than the phrases containing non-adjacent controls. Vilkaite (2016) concluded that collocational status including both adjacent and non-adjacent collocations significantly predicted the processing time when the whole phrase reading measures analysed. However, when the early measures of the final word reading times analysed, non-adjacent collocations did not facilitate the final word reading times of the nouns. In this regard, it should be noted that collocations' adjacency status affected the processing speeds differently in the early recognition processes and later integrative processes.

In addition to these behavioural research methods, a few studies have employed neuropsychological methods to look at MWS. Molinaro, Canal, Vespignani, Pesciarelli, and Cacciari (2013) examined the processing of collocational complex prepositions (e.g. *in the hands of*). These were chosen because they include a content word (e.g. *hands*), and they allow insertions (e.g. *in capable hands*). They focussed on the point that the regularity of the collocational complex prepositions could facilitate the processing of the string as a unit, but at the same time when a content word is inserted into the unit, it could make the processing of the string more difficult. Furthermore, they examined whether the string would be more difficult to process when the inserted adjective modifies the internal noun. To be able to answer these questions, they conducted an electrophysiological event related potentials (ERP) study. Adult L1 speakers of Italian participated in the ERP experiment. The authors firstly selected 56 familiar Italian collocational complex prepositions. Each collocational complex preposition was presented in two conditions: a standard condition in which collocational complex prepositions were presented in their default forms, without insertion of any additional words; and an insertion condition in which an adjective was inserted just before the nouns. The adjectives were chosen in such way as to make the collocational complex prepositions acceptable and natural in Italian. The sentences were presented word-by-word to the participants, and each word was presented for 300 ms. Molinaro et al. (2013) found that the nouns elicited a smaller N400 in the insertion than the standard conditions. This is probably because the inserted adjective restricted the range of possible continuations of the sentence at various levels. At a semantic level, adjectives modify the specific categories of nouns, and at a grammatical level, grammatically masculine singular adjectives require a masculine singular noun. Overall, Molinaro et al. (2013) showed that collocational complex prepositions can be internally modified without disrupting their processing.

Siyanova-Chanturia, Conklin, Caffara, Kaan, Van Heuven, (2017) looked at the processing of binomials (e.g. *knife and fork*) by the ERP method. More specifically, they explored the electrophysiological responses to highly predicted final words within binomials. This study included two ERP experiments. In the first experiment, participants' brain activity was recorded while they read three types of phrases: (1) frequent binomials (e.g. *knife and fork*); (2) infrequent novel phrases similar in association strength to binomials (e.g. *spoon and fork*); (3) non-associated, unattested semantic violations (e,g. *theme and fork*). In the second experiment, participants read the same stimuli without the conjunction "and" (e.g. *knife-fork* vs *knife-spoon*). The authors predicted that if a processing advantage for frequent binomial phrases over novel expressions are because of binomials' being uniquely predictable (due to their phrasal frequency), then they expected to find larger P300 amplitudes for binomials relative to novel phrases. For the first experiment, 120 matched triplets were selected from the BNC. The items used in experiment 2 were identical to experiment 1 except for the conjunction "and". The words in each item was presented one-by-one, and each individual word was presented for 300 ms. Siyanova-Chanturia et al. (2017) found that binomials elicited larger positivity (around 300 ms) than novel but strongly associated phrases. Furthermore, they found that binomials (e.g. *knife-fork*) and novel but strongly associated phrases (e.g. *knife-spoon*) elicited comparable waveforms in the P300 and N400 time windows, when phrases were presented without the conjunction "and". Siyanova-Chanturia et al. (2017) provided electrophysiological evidence in support of the view that language users are sensitive to the phrasal frequency.

In addition to the studies looking at the comprehension of MWS, it is also important to have a look at the studies focussing on the production of MWS since in many ways they investigate similar research questions adopting a slightly different perspectives from the studies focussing on the processing of MWS. Arnon and Cohen Priva (2014) investigated the effects of word and multi-word frequency counts on the phonetic duration of words in spontaneous speech. In other words, they examined whether the relationship between the word and multi-word information changes across the frequency continuum. They extracted trigrams from a corpus of conversational speech recorded from the L1 speakers of American English. They used the same corpus to calculate the word, bigram and trigram frequency of the items. They found that the effect of multiword frequency information increased for highly frequent trigrams while the effect single word frequency information decreased. That is to say, repeated usage of the MWS leads to a growing prominence of multiword information without removing the effect of single word information. However, the effect of single word frequency remained significant even for highly frequent sequences. They conclude that these findings provide empirical support for a single-system view of language at which all linguistic input is processed by a similar cognitive mechanism.

This review highlights that frequency and the transitional probabilities, also known as collocational strength, of the MWS play an important role for L1 speakers' processing and production speeds of MWS. The experimental designs using various methodological paradigms response-time, eye-movement, and ERP provided empirical evidence for L1 speakers' faster processing of MWS than control phrases (e.g. Durrant & Doherty, 2010; Arnon & Snider, 2010; Vilkaite, 2016; Siyanova-Chanturia et al. 2017). Furthermore, Vilkaite (2016) provided evidence that alongside adjacent collocations, L1 speakers process the non-adjacent collocations faster than control phrases. It should be noted that frequency, and collocational strength of the MWS are not the only factor affecting the processing of MWS, Jolsvai, et al. (2013) found that L1 speakers are sensitive to the meaningfulness of the MWS. The following section (5.2) reviews the psycholinguistic experiments which compare L1 and L2 processing of MWS.

5.2 Processing MWS in L2

A number of studies explored the influence of L1 intralexical knowledge on the processing of L2 collocations using RT based methods. Wolter and Gyllstad (2011) investigated if L2 collocations which had an equivalent form in the L1, that are also known as congruent collocations, would be processed faster than the L2 collocations which had no equivalent form in the L1, that are incongruent collocations. They also investigated if both types of collocations

(L1-L2 and L2-only) would be processed faster than random word combinations. Adult L1 Swedish-English L2 (advanced) participated in the study and L1 speakers of English served as a control group. Since it is consistent with the canonical word-order of both of Swedish and English languages, the authors preferred to use verb-noun collocations. Using the BNC, they identified two- and three-word verb-noun (object) collocations. Three types of items were constructed: lexically congruent (L1-L2) collocations, lexically incongruent collocations (L2-only) collocations, and unrelated verb-noun pairs were developed to serve as baseline items. A primed LDT task was designed to measure participants' RTs. The prime word (the verb) was presented for 250 ms. Wolter and Gyllstad (2011) found a priming effect for both collocational conditions in the responses of L1 English group. For the L1 Swedish-English L2 group, only a significant difference was found between L1-L2, lexically congruent and unrelated (baseline) items. No significant difference was found between the comparisons of L2-only, lexically incongruent, and unrelated items. It is noteworthy that their item-based analysis indicated that the results cannot be generalised to all of the items in the experiment. Wolter and Gyllstad (2011) concluded that L1 intralexical knowledge makes the L2 collocations which have an equivalent form in the L1 more readily accessible for L2 users. In other words, simultaneous spreading activation in both L1 and L2 leads to a higher level of priming. One shortcoming of this study was that the authors did not seem to control for the frequency of the individual words and collocational frequency between L1-L2 and L2-only collocational conditions so that the frequency of the individual words or a potential difference between the collocational frequency of the two conditions might have affected the results.

Wolter and Gyllstad (2013) explored how high-proficiency learners of English processed collocations under three conditions: congruent, incongruent, and non-collocational (baseline) items. Different from Wolter and Gyllstad (2011), the congruent and incongruent items were matched with range of collocational frequencies. Using the Corpus of Contemporary American English (COCA), 40 congruent and 40 incongruent collocations were selected. The selected collocations were matched for collocational frequency. In addition to these 80 collocations, 80 non-collocational items were constructed by the authors. They randomly matched mid to high frequency nouns with mid to high frequency nouns to create a list of non-collocations items (e.g. *red benefit*, *willing car*). The items in three conditions were matched for item length, noun frequency, and adjective frequency. The congruent and incongruent collocations were also matched for collocational frequency. An acceptability judgment task (AJT) was used to assess RTs. Since there is rarely anything that can be grammatically wrong with adjective-noun pairs unless the pairs mean something physically impossible, the authors used an alternate phrasing to encourage participants to have more conservative approach. They asked participants to decide if the word combinations are commonly used in English. Two groups of participants took part in the study: L1 Swedish-English L2 (advanced), and (L1) speakers of English. Wolter and Gyllstad (2013) found that L1 speakers of English group responded significantly faster than L1 Swedish-English L2 to the incongruent items. No significant differences were found in respect to the congruent and the non-collocational items. The L1 speaker participants' RTs to congruent and incongruent items were significantly faster than the non-collocational items. L1 speakers' RTs to the congruent collocations were not significantly different from the incongruent collocations. However, the L1 Swedish-English L2 participants' RTs were significantly faster to the congruent collocations than the incongruent collocations. Wolter and Gyllstad (2013) concluded that learners recognise the congruent collocations faster than the incongruent collocations. Furthermore, advanced learners are sensitive to the collocational frequency regardless of collocations' being congruent or incongruent. To account for the faster processing of congruent collocations, Wolter and Gyllstad (2013) put forth an explanation that the processing advantage might be attributable to acquisitional advantages. They are related to two closely related theories “age of acquisition” (AoA), and “onset of acquisition” (OoA). The basic idea behind these theories is that words learned earlier entrenched more than the words learned later.

Wolter and Yamashita (2015) tested the lemma-based explanation for the accelerated recognition of congruent collocations. They examined if collocations that are legitimate in the L1, but not in the L2 are still activated when processing collocations in the L2. If activation was found, it would suggest that collocational knowledge is copied into the lemma-level of the lexical entry of the L2. In turn this would provide evidence for the lemma-level explanation for accelerated processing of congruent collocations. If no activation was found, this would indicate that copying collocational knowledge into the L2 is not the likely option, in this case AoA, or OoA would provide a better explanation. Furthermore, this study aimed to improve the previous research in this field by including L2 learners of different proficiencies and both adjective-noun and verb-noun pairs in the experiment. The participants of the study were one group of English native-speakers, the advanced and intermediate L2 groups consisted of adult L1 Japanese speakers. Three types of items were constructed: English translations of collocations that were acceptable in Japanese, but not in English (J-only items, e.g. *far eye*, *buy anger*); collocations that were acceptable in English, but not in Japanese (E-only items, e.g. *low speed*, *catch breath*); non-collocational items to be used as baseline (e.g. *wet attention*). The study used a double LDT, which presented both words on the screen at the same time. Wolter and Yamashita (2015) found no evidence to support that collocational knowledge is copied into the lemma level of the lexical entry of the L2 as whole units because either L2 group did not process the J-only collocations faster than the baseline items. Another key finding was that E-only items did not elicit faster RTs for either L2 groups. Based on these findings, Wolter and Yamashita (2015) suggested the possibility that AoA/OoA may provide a better explanation for the congruency effect observed in the processing of L2 collocations. One possible limitation of this study is that the use of the double LDT could have encouraged the participants to have focus only on form, it is therefore using a task that encourages the participants to focus on meaning would be ideal (Wolter & Yamashita, 2017).

Wolter and Yamashita (2018) followed up on the studies (Wolter & Gyllstad 2011; 2013; Wolter & Yamshita 2015), investigating the effects of L1 congruency, single word frequency, collocational frequency, and L2 proficiency on L2 collocational processing. One group of English L1 speakers and two groups of L1 Japanese-English L2 speaker group took part in the experiment. The L2 groups consisted of advanced-level (L1 Japanase) L2 speakers of English and intermediate-level (L1 Japanese) L2 speakers of English. Four types of items were constructed for this study: congruent collocations; English-only (incongruent) collocations; Japanese-only (translated) collocations (e.g. *tall danger*); baseline items. Japanese-only items were created in the same way they were developed in the study by Wolter and Yamshita (2015). All items were checked against the COCA. To ensure that participants attended the meaning of the items, they used an AJT. Wolter and Yamashita (2018) found no significant differences in RTs for either L1 Japanese-English L2 speaker groups comparing the Japanese-only items with baseline items. Both L1 Japanese-English L2 speaker groups responded to congruent collocations faster than the English-only (incongruent) collocations. The L1 speakers of English responded faster to the incongruent collocations, and j-only items faster than the advanced level L1 Japanese-English L2 speaker group. The AoA/OoA effect, that is the age or order in which something is learned affects how deeply it becomes entrenched in the language system seems to explain the discrepancy between processing of congruent and incongruent collocations. Furthermore, Wolter and Yamashita (2018) found that L1 speakers of English and advanced level L1 Japanese-English L2 groups showed sensitivity to collocational frequency more than the intermediate level L1 Japanese-English L2 speaker group. They concluded as the language learners gain proficiency, they seem to shift away from relying on single word-level frequency to collocational frequency. However, both L1 and L2 speaker groups showed sensitivity both word-level and collocation frequency counts simultaneously.

Alongside the frequency and L1 congruency effects, psycholinguistic works also focussed on the factors such as the effects of transitional probabilities on the processing of MWS by L1 and L2 speakers. Ellis & Simpson-Vlach and Maynard (2008) investigated the effects frequency, transitional probabilities, and length of the phrases. Using various corpora including the MICASE and the academic spoken files from the BNC, and the LOB, 108 three-, four- and five-word MWS occurring in academic contexts were extracted. Processing and productions of these extracted MWS were examined for L1 and L2 speakers through a series of tasks. Multiple regression analyses revealed that L1 speakers' processing of MWS was affected by the transitional probabilities as measured by the MI-scores. Advanced level L2 speakers processing of MWS was affected by the phrasal frequency of the formula. It should be noted that there are a few shortcomings of this study, the sample sizes of the experiments were very small, and the confounding variables such as single word frequency counts within the MWS used in the experiments do not seem to be well controlled. Therefore, the results of these experiments should be approached cautiously.

More recently, Yi (2018) examined L1 and L2 advanced level speakers' sensitivity to collocational frequency and transitional probabilities of English adjective-noun collocations, adopting a phrasal judgment task. In addition. Yi (2018) investigated the effects of cognitive aptitudes such as working memory on the processing of collocations. As stimuli of the task, 180 adjective-noun collocations were extracted from the BNC and 180 non-collocational filler (e.g. *religious morning*) items were created. Three discrete bins were created for collocational frequency and transitional probabilities as measured by the MI-score, as high, mid and low. Statistical analyses revealed that both L1 and L2 speakers were sensitive collocation frequency and transitional probabilities of the collocations. Surprisingly, L2 speakers were even more sensitive to the collocation frequency and transitional probabilities of the collocations more than the L1 speakers. Furthermore, none of the cognitive aptitudes measured such as implicit/explicit language aptitude, working memory, moderated the L1 and L2 participants' sensitivity to collocation frequency and transitional probabilities of the items. For L1 speakers, implicit language aptitude seemed to facilitate the processing of collocations since it reduced the RTs, whereas explicit language aptitude played a negative role in the processing of the collocations. In contrast, for advanced L2 speakers, implicit language aptitude did not affect the processing of collocations, but explicit language aptitude played a facilitative role. Based on the results of Yi (2018) and Ellis et al. (2008), it is possible to say that so far empirical studies have produced conflicting findings related to the effect of transitional probabilities on the L2 processing of MWS.

The vast majority of the experimental works investigating the processing of collocations approach to the collocations through the lens of frequency-based approaches (e.g. Durrant & Doherty, 2010; Wolter & Yamashita, 2018, see also section 2.2.1 for a discussion on frequency-based approach to operationalising collocations). Gyllstad and Wolter (2016) investigated if there is a processing cost for collocations which are defined according to the phraseological tradition for L1 and L2 speakers of English. The participants of the study were L1 Swedish-English L2 speakers, and L1 speakers of English. They used a semantic judgment task to assess RTs and error rates (ERs). Participants were asked to decide if the items were meaningful and natural in English. They prepared three types of items (verb-noun pairs) 27 free combinations (e.g. *write a letter*), 27 collocations (e.g. *run a risk*), and 54 baseline items (e.g. *play fruit*). It should be noted that idiomatic expressions such as *run a risk* are classified as collocations in the phraseological tradition due to the non-compositionality of their meaning (see section 2.2.1).They used Howarth's (1996) framework for item preparation. According to Howarth's (1996) framework, free combinations are the combinations of two or more words used in their literal meaning, collocations on the other hand include words one used in its literal meaning and the other used in its special meaning. Therefore, collocations are believed to have lower degree of semantic transparency. They found no significant difference with respect to RTs for all items in the task. Both groups of participants needed more time to respond to the baseline items than to the collocations, and they needed less time to respond to the free combinations than to the collocations. Furthermore, they found that higher collocation frequency counts were associated with faster RTs. They concluded that the observed processing cost for collocations, which are defined along the lines of phraseological tradition, was because the collocations were less semantically transparent than the free combinations.

Exploring an alternative methodology to examine the processing of MWS by L1 and L2 speakers, researchers also looked at their eye-movements. Siyanova-Chanturia, Conklin, Van Heuven (2011) investigated if L1 and proficient L2 speakers are sensitive to phrasal frequency during comprehension. To investigate this, they used three-word binomials (e.g. *bride and groom*). They used the BNC to choose 42 binomial expressions. To use as control phrases, they reversed the binomial expressions (e.g. *groom and bride* instead of *bride and groom*). Therefore, the items in two conditions were matched for single-word frequency, and length, but they were different in phrasal frequency. L1 English speakers, and L2 learners of English with high-low proficiency took part in the study. They used three eye-movement measures; first-pass reading times, total reading times, and fixation count to assess the processing of binomials. The results of all three measures revealed that L1 speakers and proficient L2 speakers read the binomials and controls significantly faster than the less proficient L2 speakers of English. Furthermore, phrasal frequency of the binomials significantly predicted the all three measures of eye-movements for both groups. The frequency counts of the individual words did not significantly predict the reading times. They concluded that both L1 and L2 speakers of English are sensitive to the frequency counts of the binomials. They are also sensitive to whether a phrase occurs in a particular configuration, which highlights the contribution of entrenchment of phrases. It should be noted that Siyanova-Chanturia, et al. (2011) provided empirical support for prediction effect, more experienced speakers of English would expect to see word *bride* after they saw the word *groom*, but they would not expect to see *bride* after they saw *groom*. Processing difference between binomials and reversed could be likely to be because of the predictability effect.

Sonbul (2015) also looked at eye-movements to explore whether L1 and L2 speakers of English are sensitive to collocational frequency of the adjective-noun colocations, and whether sensitivity to frequency is affected by L2 speakers' level of proficiency. L1 speakers, advanced level L2 learners of English participated in the study. Using the BNC, she chose adjective-noun collocations representing three levels of collocational frequency; non-collocate (e.g. *extreme mistake*), low-frequency (e.g. *awful mistake*), and high-frequency (e.g. *fatal mistake*). Each noun node was matched with two collocates, and one non-collocates. The items were put into neutral sentence contexts (e.g. *The engineer made one fatal mistake*). The stimuli included 60 adjective-noun collocations, and 30 non-collocational items. As eye-movement measures, she used first pass reading time, total reading time, and fixation count. She found that both L1 speakers and L2 learners are sensitive to the collocational frequency according to the first-pass reading time measure. According to the total reading time and fixation count measures, collocational frequency did not affect the later integrative processes. The results of Sonbul's (2015) study, and Siyanova-Chanturia et al. (2011) found different results with respect to phrasal frequency effects on early and late measures. Sonbul (2015) argued that the conflicting results between these two studies can be attributed to the differences between binomials and collocations. Both L1 and L2 speakers are able to recover from processing an infrequent but plausible collocation, but reading reversed binomials should be a more difficult processing experiences.

A wealth of studies both on L1 and L2 processing of MWS have demonstrated that MWS are processed quantitatively faster than control phrases (e.g. Arnon & Snider, 2010; Cangir et al. 2017; McDonald & Shillcock 2004; Siyanova-Chanturia et al. 2011; Wolter & Yamashita, 2015; 2017; Wolter & Gyllstad 2011; 2013). The studies using various methodological paradigms produced empirical evidence that language users are sensitive to frequencies of the phrases. In connection with this, these studies played an important role in providing empirical evidence against the traditional distinction between and grammar and lexicon. However, one weakness of the studies in this field is that they focused predominantly on a few selected languages, especially on English. As mentioned previously in Part 1 and Part 2, a potential problem with this approach is that the status of MWS as a general feature of language might not be sufficiently established. Therefore, experimental works need to focus on MWS in typologically different languages. This thesis addresses this gap from a psycholinguistic perspective by comparing the on-line processing of MWS in an agglutinating language Turkish, and a non-agglutinating English. More specifically, the psycholinguistic experiments reported in chapter 6 explores the extent to which L1 English and Turkish participants rely on the same mechanisms, single-word and collocational frequency counts for processing adjective-noun collocations. The present study also investigates whether the participants' response times (RTs) for adjective-noun collocations mirror the patterns emerged from the corpus study, in relation to the collocations' frequency of occurrence and collocational strength. It should be noted that lemmatieed collocations have considerably higher association scores in Turkish than their translation equivalents in English, as the findings of the corpus study showed (see sections 4.3.1 for a detailed discussion).

Chapter 6: Processing Adjective-noun Collocations in English and Turkish

The experiments reported in this chapter looks at the processing of adjective-noun collocations in Turkish and English. The first aim of this study is to explore whether the same variables affect the processing of Turkish and English adjective-noun collocations by L1 speakers of these languages. In order to explore this, this study closely examines the influence of both single word, and collocational frequency counts on the processing of collocations in Turkish and English by L1 speakers of these languages. The second aim of this study is to explore the variables which affects the processing of adjective-noun collocations by L1 Turkish-English L2 advanced speakers in English. To test this, the current study closely examines the influence of the factors single word and collocational frequency counts on the processing of collocations by L1 Turkish-English L2 advanced speakers in English. Through this, it will be possible to gain a clearer understanding of the role of the single word and collocational frequency counts in collocational processing in L1 and L2, and whether they have different effects in English and Turkish due to the typological difference of these languages. The collocations are carefully operationalised according to the frequency-based approach to collocations (Evert, 2008, Gablasova et al. 2017, Hunston, 2002), which draws on quantitative evidence on word co-occurrence in corpora. Adjective-noun collocations are used to explore the processing of collocations in English and Turkish. The main reason for choosing adjective-noun collocations is that they occur in the same syntactic order in which adjectives precede the nouns in both Turkish and English, hence they should be fully comparable for both strength and directions of the syntagmatic associations. This study utilises adjective-noun collocations at different levels of collocational frequency and collocational strength, which are classified as high-frequency, low-frequency, and non-collocational (baseline) items. The reason for using collocations at two different levels of collocational frequency and strength is that this allows me to compare how quickly the collocations at different levels of frequency and associations are processed in English and Turkish. Specifically, this study sought to test the following hypotheses.

1. L1 (native) speakers of both Turkish and English will process the high-frequency collocations faster than the low-frequency collocations, which will in turn would be processed faster than the baseline items. This hypothesis is based largely on the results of the previous studies that L1 speakers and advanced L2 speakers which show sensitivity to collocational and phrasal frequency (see Arnon & Snider, 2010; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018; and Siyanova et al. 2011).
2. L1 (native) speakers of Turkish will process both high-frequency and low-frequency adjective-noun collocations more rapidly than L1 (native) speakers of English. This hypothesis is based on results of a corpus study reported in Chapter 4 (see 4.3.1 for summary and discussion of the findings). Lemmatised adjective-noun collocations have considerably higher association scores in Turkish than their translation equivalents in English.
3. Lemmatised collocation-level frequency counts will have a larger effect on the processing of Turkish adjective-noun collocations than English adjective-noun collocations. This hypothesis is based on results of a corpus-based study conducted as part of this PhD thesis (see Section 4.3.1). On average, lemmatised adjective-noun collocations have a considerably higher frequency than their equivalents in English.
4. Word-level frequency counts of the nouns will have a larger effect on the processing of Turkish adjective-noun collocations than English adjective-noun collocations. This hypothesis is based on the results of the corpus study (see section 4.3.1). Turkish adjective-noun collocations have many more inflected forms than their equivalents

English. Therefore, the speakers of Turkish are more likely to be sensitive to the noun frequency counts.

5. L1 Turkish-English L2 speakers (advanced) will show sensitivity to both collocational and single-word level frequency counts. This hypothesis is based on the result of previous psycholinguistic experiment by Wolter and Yasmashita (2017), that the advanced level L2 speakers will attend to both single word and collocational frequency counts.

6.1 Pilot studies

Since no study has examined whether collocations are processed differently in typologically different languages using response times methods, it was necessary to pilot the items with different response-times-based tasks to see which technique has a potential to provide fruitful data for this study. In this section, some of these pilot tests are reported. I firstly piloted the same stimuli with using LDT under different stimulus onset asyncrony conditions (SOA) to find out the suitable time interval between onset of the prime and the onset of the target. I firstly piloted the LDT with 50 ms SOA under masked priming condition. Five NSs of English completed the task. The participants were not told the purpose of the experiment. They were instructed that they will see string of letters on the screen one after another and asked to decide whether they are words or non-words in English, by pressing on specified yes and no buttons using Logitech Dual Action Game Pad. The participants were undergraduate and postgraduate students at Lancaster University, and were 20-31 years of age (M=23.8, SD=3.96). The results showed that RTs were faster for the low-frequency (low-frequency: 503 ms (SD =100)) and high-frequency collocations (high-frequency: 504 ms (SD =100), compared to the baseline items (baseline = 511 (SD =103)).

The findings of the first pilot test showed that participants' RTs for the collocations in the high-frequency and low-frequency conditions received shorter RTs than the items in the baseline condition. It is not surprising to see that English L1 speaker participants are possibly more sensitive to adjective-noun collocations than the random combinations of the adjectives and nouns in the baseline condition. At the same time, however, it was surprising to see that participants' RTs for the collocations in the low-frequency and high-frequency conditions were not considerably different from each other. This was an unexpected finding since L1 speakers were expected to process the high-frequency collocations faster than the low-frequency collocations.

I secondly piloted an LDT with 100 ms SOA with the same items under unmasked priming condition. Seven NSs of English completed the task. The participants were 7 undergraduate students at Lancaster University, and were 19-21 years of age (M=20.28, SD=0.69). The results showed that RTs were faster for the high-frequency collocations (high-frequency: 564 ms (SD =117)) compared to the low-frequency (low-frequency: 580 ms (SD =176), and the baseline items (baseline = 579 (SD =167)). The findings showed that participants' RTs for the collocations in the high-frequency condition was shorter than RTs for the collocations in the low-frequency and the baseline conditions. It is not surprising to see that English L1 speakers are more sensitive to high-frequency collocations than the low-frequency collocations in and the items in the baseline condition. It was however unexpected that participants' RTs for the baseline items was shorter than RTs for the low-frequency collocations.

I thirdly piloted an AJT with the same items. The participants' RTs were faster for the high-frequency collocations (high-frequency: 864 ms (SD =325)) compared to the low-frequency (low-frequency: 1032 ms (SD =423)), and the baseline items (baseline = 1254 (SD =511)). The fidnings were not surprising. Based on this finding it was possible to say that L1 speakers were sensitive to the frequency of the collocations and the AJT was used in this study. Furthermore, some other recently published studies investigating collocatinal processing used AJTs (Wolter & Yamshita, 2018; Yi, 2018). On this basis, I chose to employ AJTs to explore Turkish and English speakers' processing of collocations in the two languages.

6.2 Method

6.2.1 Participants.

The participants were one group of L1 (native) English speakers, and two groups of L1 Turkish-English L2 speakers. The L1 English group consisted of 25 undergraduate and 6 postgraduate students all from a university in the UK (n=31). The first group of L1 Turkish-English L2 participants consisted of 22 undergraduate and 10 postgraduate students, they were all from two different universities in Turkey (n=32). They were identified as an advanced level learners of English and given the AJT to complete in English. The other group of L1 Turkish-English L2 participants was comprised of 40 undergraduate and 6 postgraduate students (n=46), all from two different universities in Turkey. They were identified as intermediate level learners of English and given the AJT in Turkish. The LexTALE, a test of vocabulary knowledge for advanced learners of English, was used to identify the advanced level L1 Turkish-L2 English speakers and consequently give them to the AJT in English. Therefore, all of the participants (including L1 English, L1 Turkish-English L2 groups) were administered the LexTALE (Lemhöfer & Broersma, 2012) to assess their English vocabulary knowledge as a proxy for general English proficiency. The LexTALE was chosen because it is a quick, validated, and practically feasible test for identifying advanced level L2 speakers of English (Lemhöfer & Broersma, 2012). It consists of a simple, un-speeded, visual LDT. To identify the

L1 Turkish-L2 English advanced learners, a cut-off LexTALE score was determined. Lemhöfer & Broersma, (2012) reported that on average, a Quick Placement Test score of 80% corresponds to a LexTALE score of 80.5%. Therefore, a LexTALE score of 80.5% was used as a cut-off score to allocate participants the AJT in English. The participants who had scores below 80.5% were allocated the AJT in Turkish. On average the advanced group had a significantly larger vocabulary size than the non-advanced group (84.85 vs 62.52, $t_{(74.873)} = 16.27$, $p < 0.05$), and the L1 English group had a significantly larger vocabulary size than the advanced group (90.82 vs 84.85, $t_{(56.072)} = 5.15$, $p < 0.05$).

In addition to the LexTALE test, all participants answered a questionnaire that included questions about some personal information (age, gender, dexterity, and visual acuity). One participant reported a problem about his natural/corrected vision. Therefore, his data was removed. The L1 Turkish-English L2 participants were also asked to provide information about their English learning background including self-reported proficiency in English, starting age of learning English, length of studying English through formal education, and length of residence in an English speaking country. Among the L1 Turkish-L2 English groups there was no significant difference in the mean starting age of English learning between the advanced and non-advanced (10.96 years old) and the (11.96 years old) groups ($t_{(55.92)} = 0.73$, $p = 0.46$). The advanced group had studied English for a significantly longer period than the non-advanced group (14.81 vs 12.71 years: $t_{(55.219)} = 2.46$, $p < 0.05$). Twenty one participants in the advanced group had lived in an English speaking country for longer than one month, whereas nine participants in the intermediate group had lived in an English speaking country for longer than one month. Table 6.2 summarises the participants' background information.

Table 6.2 Biographical data for participants (standard deviations are in parentheses)

Group	N	Age	Dexterity (R/L/Both)	Gender (M/F)	Self-reported proficiency scores[a] Speaking	Listening	Reading	Writing	LexTALE in English
L1-English (AJT in English)	31	20.58 (2.16)	28/3/0	13/17[b]	-	-	-	-	90.82 (3.71)
L1-Turkish (AJT in English)	32	24.43 (4.01)	28/3/1	16/14	5.28 (0.44)	5.62 (0.48)	5.65 (0.47)	5.53 (0.49)	84.85 (5.22)
L1-Turkish (AJT in Turkish)	46	26.5 (5.51)	39/7/0	19/27	4.82 (0.73)	4.97 (0.73)	5.3 (0.58)	5.1 (0.66)	62.52 (6.7)

Note: LexTALE = Lexical test for advanced learners of English.

1[a] = Beginner, 6 = Very advanced

[b]Excluding one participant who did not indicate gender.

6.2.2 Item development.

All the English items were extracted using the British National Corpus (BNC), and all the Turkish items were extracted using the Turkish National Corpus (TNC). The items in both Turkish and English fell into one of three critical conditions: (1) high-frequency collocations, (2) low-frequency collocations, (3) non-collocational (baseline) items. The non-collocational (baseline) items were used for establishing a threshold reaction time (RT) for measuring the relative RTs for the items in conditions (1) and (2), and to ensure that participants did not develop a familiarity effect for the task. Single word frequency counts, collocation frequency counts, and LD scores of the English items were obtained from the BNC and Turkish items from the TNC. The non-lemmatised frequency counts at both single word and collocational levels were used in the item development. As reported previously (see section 4.3.1), the corpus study found no considerably large differences of frequency between the lemmatised and unlemmatised forms of the English collocations, however it found larger differences between the two forms of the Turkish collocations. All single word and collocational frequency counts were log transformed using SUBTLEX Zipf scale (Van Heuven, Mandera, Keuleers, and Brysbaert, 2014). It is a logarithmic scale (from 1 to 7).

To distribute the items between (1), (2), and (3), a LD measure was specifically preferred. The LD is a standardised measure of collocational strength with a maximum value of 14, making it a comparable measure across corpora (Gablasova et al. 2017). The LD measure is not negatively link to high-frequency of occurrence, unlike MI scores (Gablasova et al. 2017; Bartsch & Evert, 2014), and thus it is possible to extract high-frequency collocations using LD measure. Through an initial corpus study (see section 4.2.2), I explored the scales of relative collocate frequency (RCF) and LD scores of adjective-noun collocations in English and Turkish. According to the scale of LD scores presented in Section 4.2.2, adjective-noun collocations in both English and Turkish with LD scores of ≥7 were defined as high-frequency collocations within a 3-3 collocation window span. Adjective-noun collocations with an LD score of between 2 and 4 were defined as low-frequency collocations within a 3-3 collocation window span. Finally, the items which had negative LD scores were defined as non-collocational items because the LD scores become negative when the co-occurrence frequency is lower than what would be expected by chance.

To select high-frequency English collocations, the nouns in the BNC word frequency list were checked for whether they collocate with an adjective in a way that meets the LD cut-off scores determined by the corpus-based study presented in Section 4.2.2 for high-frequency collocations. An initial list of 36 collocations satisfied the selection criteria for high-frequency collocations: an LD-score of ≥7, and within a 3-3 collocation span, (the BNC XML edition). To satisfy the LD cut-off scores for high-frequency collocations, predominantly high frequency nouns were checked to ensure that all collocations would be known by the L1 Turkish L2 English participants. Four of the collocations in the list were discarded from the study, because they were incongruent with Turkish (e.g *supreme court*, *british library*), considering the empirical evidence that lexical congruency affects collocational processing in L2 (Wolter & Gyllstad, 2011, 2013). Since one of the goals of this study is to investigate the extent to which collocation and single-word frequency affect the processing of collocations by L1 Turkish-English L2 speakers including incongruent collocations would be a confounding variable. Two collocations were also discarded because their component words were cognates for Turkish (*modern art*, *high standard*), cognates are potentially concern for eliciting faster reaction times (Lemhöfer et al. 2008). Eventually, a list of 30 high-frequency English collocations remained. The mean LD score for all high-frequency collocations was 7.80, with a low score of 7.0 (for the items *Dark hair* and *Left hand*), and with a high score of 10.95 (for the item *prime minister*).

To select low-frequency English collocations, the nouns in the BNC word frequency list, (which had not been not used) for the high-frequency collocations were checked for whether they collocate with an adjective in a way that meets the LD cut-off scores for the low-frequency collocations. The selected low-frequency collocations had LD scores of 2 and 4 within a 3-3 collocation window span to satisfy the criteria established in Section 4.2.2. The low-frequency collocations were also legitimate word pairs in English, occurring in the BNC. It was therefore confirmed that the selected low-frequency collocations were corpus-verifiable items and their collocational frequency was considerably lower than the high-frequency collocations. As with the high-frequency collocations, the low-frequency collocations were also congruent with Turkish. It should be noted that none of the nouns and adjectives used for the items in the high-frequency collocations were not used for the items in the low-frequency collocations, however single words (both adjectives and nouns) in both types of items were closely matched for length, and frequency. A list of 30 low-frequency collocations were extracted. The mean LD score for all the low-frequency collocations was 3.24, with a low score of 2.54 (for the item *away game*, and with a high score of 3.91 for the item *Vital information*). The high-frequency and low-frequency English collocations were closely matched for abstractness.

The noncollocational baseline items consisted of random combinations of the nouns used for the high-low-frequency collocations with adjectives which had not been used previously. For instance, *european community* is a high-frequency collocation, and *fair idea* is a low-frequency collocation, the same nouns (*community* and *idea*) were randomly matched with the adjectives *short* and *cold*, which had not been previously used for the high and low-frequency collocations, to produce the baseline items, *short community*, and *cold idea*. All combined nouns and adjectives for the baseline items were checked against the BNC to make sure that there was no occurrence. If any occurrence was found in the BNC, the LD-scores were checked to make sure that they were negative values. If the combinations produced positive LD-scores, the process was repeated. I eventually obtained a list of 60 baseline items. Nonetheless, given the very large size of the BNC, it was not possible to fully eradicate the positive LD scores. I therefore decided to retain two items with positive but very low LD scores. These items included *clear trade* (LD=0.45), and *public class* (LD=0.16). The mean LD score for all baseline items was -0.93, with a low score of -3.22 (for the item *dirty time*) and with a high score of 0.45 (for the item *clear trade*). Furthermore, to make sure that the baseline items are not meaningful, I conducted a very small-scale norming study. Seven L1 speakers of English from the same population as the participants of the main study took part. The participants were asked to rate the naturalness of the items including high-frequency, low-frequency and baseline from one to ten. I determined naturalness rating of 3 as a threshold to exclude the items from the list. None of the baseline items reached naturalness rating of 3.

On the one hand, repeating the same nouns in different conditions was an ideal way of ensuring that the single word length and frequency of the nouns in the collocational and baseline conditions were perfectly matched, both of which are characteristics that are known to affect processing speed. On the other, this meant that each noun appeared in the task twice, once in the collocational conditions and once in the non-collocational condition. This inevitably introduced another potential confounder that it possibly lowered the activation thresholds for the nouns that had been seen in a different condition by the participants. To address this, all items were presented to the participants in an individually randomised order. Thus, any advantage gained from a seeing word for a second time was evened out both within the individual participant's test and across all of the participants as whole. The adjectives chosen for the baseline items were closely matched with the adjectives used for the high-frequency and low-frequency collocations in terms of single word length and frequency.

Table 6.3 Summary of English items (Log transformed frequency counts and standard deviations are in parentheses)

Item type	High-frequency collocations	Low-frequency collocations	Non-collocations	Statistical comparison
Item length	10.86 (2.97)	11.1 (2.3)	11.1 (2.52)	*W*=401, *p*=.46
Adjective frequency	5.17 (0.31)	5.17 (0.42)	5.15 (0.24)	*W*=467.5, *p*=.79
Noun frequency	5.36 (0.29)	5.36 (0.21)	5.36 (0.25)	*W*=415.5, *p*=.60
Collocational frequency	4.03 (0.34)	2.7 (0.3)	1.18 (0.52)	*W*=891, p<.05
Log Dice scores	7.8 (0.82)	3.24 (0.39)	-0.93 (0.85)	

Note: The statistical comparison is based on the Wilcoxon test comparing medians between high-frequency and low-frequency collocations.

The same procedure and criteria used for extracting high and low-frequency English collocations was followed to extract Turkish collocations. Firstly the nouns in the TNC word frequency list were checked for whether they collocate with an adjective in a way that meets the LD-score based criteria for high-frequency collocations. The selected high-frequency Turkish collocations had LD score of ≥7. A list of 26 high-frequency Turkish collocations were obtained. The mean LD score for all high-frequency collocations was 7.92, with a low score of 7.0 for the item *bilimsel araştrma* (the translation equivalent is *scientific research*) and with a high score of 9.84 for the item *sosyal güvenlik* (the translation equivalent is *social security*). To select the low-frequency Turkish collocations, the nouns in the TNC word frequency list, (which had not been used for the high-frequency collocations) were checked for whether they collocate with an adjective in a way that meets the LD-score based criteria for the low-frequency collocations. The selected low-frequency Turkish collocations had LD scores of between 2 and 4 within a 3-3 collocation window span. The words (both adjectives and nouns) in the both high and low-frequency collocations were closely matched for single word length, and frequency. A list of 26 low-frequency collocations were extracted. The mean LD score for all low-frequency collocations was 3.6 with a low score of 2.49 for the item *nitelikli işçi* and (the translation equivalent is *qualified worker*) and with a high score of 4 for the item *acı haber* (the translation equivalent is *sad news*). The high and low-frequency Turkish collocations were closely matched for single word frequency, length and abstractness.

To produce the Turkish noncollocational items, the procedure used to extract English noncollocational items was repeated. That is, the nouns used for the high and low-frequency collocations were randomly matched with a list of adjectives which have not been used for the high and low-frequency collocations. Since the same nouns were repeated both in collocational and noncollocational (baseline) conditions, the single word length and frequency of the nouns in the collocational and baseline conditions were perfectly matched. The adjectives chosen for baseline items were closely matched with the adjectives used for the high and low-frequency collocations in terms of single word length and frequency. All combined nouns and adjectives used for the baseline items were checked against the TNC to make sure that there was no occurrence. If any occurrence was found in the TNC, the LD scores were checked to make sure that they were negative values. Since the TNC is considerably smaller than the BNC, it was possible to fully eradicate the positive LD scores. All the baseline items except the item *gerçek yıl* (the translation equivalent is *real year*) had no attestation in the TNC. The item *gerçek yıl* has a negative LD score of -1.79. Eventually, a list of 52 baseline items were obtained.

Table 6.4 Summary of Turkish items (Log transformed frequency counts and standard deviations are in parentheses)

Item type	High-frequency collocations	Low-frequency collocations	Non-collocations	Statistical comparison
Item length	10.65 (2.18)	10.65 (1.59)	10.42 (1.81)	W=314.5, p=.66
Adjective frequency	5.39 (0.29)	5.33 (0.4)	5.29 (0.31)	W=347.5, p=.86
Noun frequency	5.32 (0.42)	5.33 (0.17)	5.33 (0.32)	W=292.5, p=.40
Collocational frequency	4.03 (0.4)	2.77 (0.16)	0.03 (0.21)	W=676, p<.05
Log Dice scores	7.92 (0.77)	3.6 (0.26)	-0.03 (0.024)	

Note: The statistical comparison is based on the Wilcoxon test comparing medians between high-frequency and low-frequency collocations.

6.2.3 Procedure.

AJTs were used to assess RTs, on high-frequency, and low-frequency collocations, and baseline items in English and Turkish. The task was administered using PsychoPy software (Peirce, 2007). AJTs ask the participants to simply indicate whether or not the items are acceptable. It has most frequently been used with grammatical acceptability in which judgments are more straightforward. However, the vast majority of the adjective-noun combinations are mostly grammatical unless the combinations of words indicates something that is highly unlikely (e.g. *old child*). Therefore, adjective-noun combinations can be perceived as acceptable if some flexibility is used in interpreting them. To avoid this obstacle, I followed the alternate phrasing used by Wolter & Gyllstad, (2013), and asked participants to indicate whether or not the word combinations were commonly used in English or Turkish, depending on the language of the experiment. The exact instructions were as follows:

In this experiment, you will be presented with 120 word combinations. Your task is to decide, as quickly and accurately as possible whether the word combinations are commonly used in English or not. For instance, the word combination *harsh words*, is a commonly used word combination in English, but *complex force* is not a commonly used word combination in English. Please press the “YES” button on the game pad if the word combination is commonly used, and “NO” button if it is not commonly used in English.

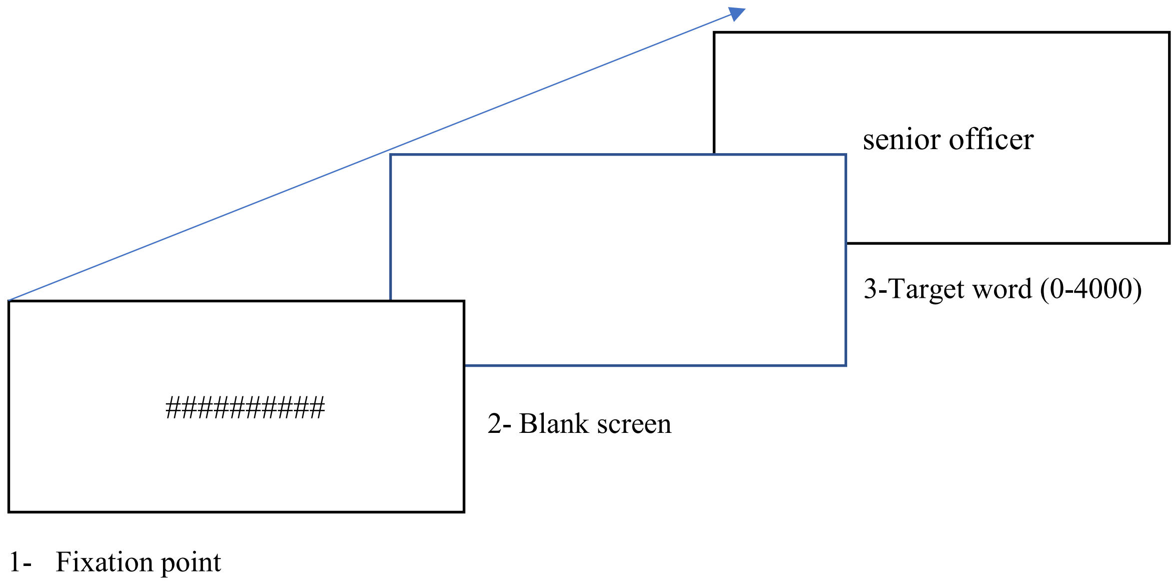

Figure 6.1 Presentation sequence for items in the AJT.

No other instructions were provided with respect to the items. The presentation sequence is shown in Figure 1. Firstly, the eye fixation (#########) was presented for 250ms, and followed by a blank screen. After the blank screen, the item was presented in lowercase in Times News Roman 12. The item remained on the screen either until the participant (via pressing a button) or after a 4000ms timeout. The presentation sequence was exactly the same for the AJT in both English and Turkish. All items were presented in an individually randomised order. To being with, I randomised the items using excel's randomise function and then PsychoPy randomised it for each participants. The participants were instructed to indicate their responses as quickly, and accurately as possible. They entered their responses using a Logitech Dual Action Gamepad. They answered YES by pressing the button corresponding to the forefinger of the dominant hand, NO by pressing the button corresponding to the forefinger of the nondominant hand. The AJT began with a practice session to familiarise the participants with the actual task. The practice session included 10 collocations and 10 baseline items which were not used in the actual task. The participants were instructed to decide whether they are commonly used word combinations in English or Turkish. The participants were allowed a short break after the practice session. Most participants completed the AJT in 5-6 minutes irrespective of the language.

6.3 Results

The main concern of the study was how participants processed the high and low-frequency collocations they viewed as acceptable, compared to the baseline items they viewed as unacceptable. In other words, I analysed the RTs to the high- and low-frequency collocations that received a "yes" response, and to the baseline items that received a "no" response. This approach could have been proven potentially problematic in two ways. Firstly either if the majority of the high- and low-frequency collocations received a "no" response; or a majority of the baseline items that received a "yes" response. Fortunately neither was the case. The L1-Turkish (AJT in Turkish) group judged 98.24% of the high-frequency collocations and 88.1% of the low-frequency collocations to be acceptable in Turkish, and they judged 81.27% of the baseline items to be unacceptable in Turkish. 21 items Turkish in total items were excluded because they did not receive any response. The L1-English group judged 98% of the high-frequency collocations, and 78.11% of the low-frequency collocations to be acceptable in English, and they decided that 78.77% of the baseline items are unacceptable in English. 12 English items were excluded because of no response. L1 Turkish-English L2 (advanced) participants judged 97.5% of the high-frequency collocations, and 76.56% of the low-frequency collocations to be acceptable in English. They decided that 71.19% of the baseline items are unacceptable in English. 17 items were excluded because they did not receive any response by L2 English participants. The second reason this approach could have been problematic is that the corpus data do not fully represent the individual experiences of the participants (see also e.g., Gablasova et al. 2017a; Gonzalez Fernandez & Schmitt, 2015). The individual differences in language experiences might have led some participants to judge some of the items based on their own language experiences of English or Turkish, which are different from the corpus-based evidence. It is also possible to say that some participants may have judged some of the baseline items based not solely on their experiences of language, but also on the plausibility of the item, which is how individual words contribute to the meaning of the item.

6.3.1 Model development.

For the statistical analysis, the lme4 package was used (Bates, Maechler, Bolker & Walker, 2014) in the R statistical software platform (R Core Team, 2016). The data were analysed using linear mixed effect modelling. The models were built using the RT as the response variable. Four models were built in total. Model 1 was built to examine the RTs in English and Turkish by L1 speakers of these languages. It included only the predictors item type (high-frequency, low-frequency, baseline), language (English and Turkish), noun frequency, adjective frequency and length - test hypotheses 1 and 2. Model 2 was built to examine the RTs in Turkish by L1 speakers of Turkish, to focus on the language specific factors predicting the RTs for Turkish collocations. Model 3 was built to examine the RTs in English by L1 speakers of English, to focus on the language specific factors predicting the RTs for English collocations. Model 4 was built to examine the RTs in English by both L1 speakers of English, and L1 Turkish-English L2 advanced speakers. In order to ensure that the effects hold beyond the items and subjects used in this experiment, all models followed the criteria of the "maximal random effect structure" justified by Barr et al. (2013). Using the MuMIn package in R (Barton, 2016), I calculated the effect sizes for all four models. The procedure generates two different R^2 values for the fitted mixed effect models: marginal and conditional. Marginal R^2 values determine the effect sizes only for the fixed effects while conditional R^2 values are associated with both the fixed and random effects. The lmerTest package in R (Kuznetsova, Brockhoff, and Christensen, 2015) was used to compute the *p*-values for each predictor variables. I also conducted likelihood ratio tests to compare versions of the models with both the main effects alone and then both the main effects and the interactions together. Following the minimal data trimming choice by Gyllstad & Wolter, (2016), and Wolter & Yamashita (2018), only the responses that were faster than 450 ms, and the responses that timed out at 4.000 ms were excluded. The remaining RTs were log-transformed using the natural logs to control for skewing. All continuous predictor variables were centred and standardised prior to the analyses. The first versus second occurrence of the nouns were entered as categorical variables.

As predicted, Turkish L1 participants' RTs were faster for the high-frequency collocations (high-frequency: 874 ms (SD =335)) compared to the low-frequency (low-frequency: 1043 ms (SD =456)), and the baseline items (baseline = 1243 (SD =506)). The same pattern was also observed in English L1 participants' RTs. They were faster for the high-frequency collocations (high-frequency: 892 ms (SD =338)) compared to the low-frequency collocations (low-frequency: 1075 ms (SD =431)), and the baseline items (baseline = 1303 (SD =527)). Model 1 was constructed to test if this finding can hold beyond the items and subjects used in this experiment, and if there is an interaction between the predictors, item types (high-frequency, low-frequency, baseline) and languages (English and Turkish). This model had by-subject adjustments to the intercept as well as by-subject adjustments to item type. That is to say, I specified subject and item as random intercepts as well as a by-subject random slope for item type. Since language is a categorical variable the effect coding was used for this model. The model included item types, language, noun frequency, adjective frequency and item length as main effects. In addition, it included item type by language as an interaction. After fitting the model, I visually inspected the residuals vs fitted plot (see appendix H), which confirmed that residuals were normally distributed.

As can be seen in Table 6.5, L1 speakers of both English and Turkish participants responded to the high-frequency collocations faster than the (non-collocational) baseline items (β_1 = -.365, [SE= .020], p < .001). They responded to the low-frequency collocations faster than the baseline items (β_2 = -.188, [SE= .020], p < .001). Re-levelling of the model from the baseline items to the weak collocations showed that L1 speakers of both English and Turkish participants responded to the high-frequency collocations faster than the low-frequency collocations (β= -.177, [SE= .018], p < .001). That is to say, L1 speakers of both English and Turkish responded faster to the collocations which occur more frequently and which are more strongly associated than to the collocations which occur less frequently and less strongly associated in their native languages. The effect of item type was significant when controlling for adjective and noun frequency counts, and item length. Language (English vs Turkish) did not predict log RTs (β_3 = -.032, [SE= .046], p > 0.1) as a main effect. Noun frequency did not significantly predict log RTs, but it should be noted that it nearly reached the level of significance (β_4 = -.038, [SE= .022], p > 0.5). Adjective frequency counts did not predict log

RTs (β_5 = .009, [SE= .007], p > 0.1). Item length significantly predicted log RTs (β_5 = .028, [SE= .006], p <.001). This indicates that as the number of letters increased in the items, participants needed a longer time to respond in both languages. I ran a likelihood ratio test to compare two versions of model (1), one with the main effects alone and the other including both the main effects and the interactions. The model comparison showed that there was not a statistically significant difference between the two models (chi-square = 0.3, p > 0.1).

Table 6.5 Selected mixed-effect model 1 comparing L1 English and L1 Turkish groups for item types (high-frequency, low-frequency and baseline)

Fixed effects	Estimate	Std. error	Df	*T*-value	*P*-value
Intercept	**.402**	**.128**	**235**	**3.14**	**<.01**
High-frequency coll.	**-.365**	**.020**	**157**	**-17.92**	**<.001**
Low-frequency coll.	**-.188**	**.020**	**158**	**-9.35**	**<.001**
English	-.032	.046	98	-0.68	>0.4
Noun Frequency	-.038	.02	222	-1.67	>.05
Adjective Frequency	.009	.007	224	1.38	>0.1
Item length	**.02**	**.006**	**221**	**4.48**	**<.001**
High-freq coll. * English	.021	.039	162	0.54	>0.5
Low-freq coll. * English	.008	.039	162	0.22	>0.8

R^2 Marginal = .17, R^2 conditional = .41

As can be seen in Table 6.6, the interactions between item types and languages did not appear to significantly predict the log RTs. Nevertheless, to be able to address the hypotheses 1 and 2 more adequately, I conducted a post-hoc test using the least square means through the emmeans package in R (Lenth, 2018), with Tukey adjustments for multiple comparisons. Specifically, the post-hoc test first investigated the log RTs for each item types (high-frequency, low-frequency, baseline) within English and Turkish. I also compared English and Turkish to detect if there is any significant difference in the log RTs for each collocational conditions (high-frequency and low-frequency). As shown in Table 6.6, L1 speakers of English needed less time to respond to the low-frequency collocations than to the baseline items (β_1 = -.192, [SE= .028], p > .001), and they needed less time to respond to the high-frequency collocations than to the low-frequency collocations (β_3 = -.182, [SE= .026], p > .001). L1 Turkish participants' log RTs followed the same pattern. They needed less time to respond to the low-frequency collocations than to the baseline items (β_4 = -.183, [SE= .026], p > .001), and they needed less time to respond to the high-frequency collocations than to the low-frequency collocations (β_6 = -.169, [SE= .025], p > .001). These findings confirmed the results of the model (1) about the item type as a significant main effect. The post-hoc test also showed that there was no significant differences in the participants' responses to the any of the item types when English and Turkish were compared. The high-frequency collocations were not responded to any faster in English by English (L1) participants than they were responded to in Turkish by Turkish (L1) speakers (β = .010, [SE= .041], p > 0.1), neither were the low-frequency collocations responded to faster in English by English (L1) participants than they were responded to in Turkish by Turkish (L1) speaker (β = .023, [SE= .043], p > 0.1).

Table 6.6 Results of post-hoc test of RTs for high-, low-frequency, and baseline items

Contrasts	Language	Estimate	Std. error	z. ratio	*P*-value
Low-freq coll. - Baseline	English	-.192	.028	-6.65	<.001
High-freq coll. - Baseline	English	-.375	.029	-12.82	<.001
High-freq - Low-freq	English	-.182	.026	-7.00	<.001
Low-freq coll. - Baseline	Turkish	-.183	.026	-6.88	<.001
High-freq coll. - Baseline	Turkish	-.353	.027	-13.05	<.001
High-freq - Low-freq	Turkish	-.169	.025	-6.57	<.001

Model 2 was constructed to predict the log RTs in Turkish by L1 speakers of Turkish only. This model had by-subject adjustments to the intercept as well as by-subject adjustments to item type. Specifically, I adjusted subject and item as random intercepts as well as a by-subject random slope for item type. Model 2 included the item types (high-frequency, low-frequency, baseline), single word frequency counts for adjectives, non-lemmatised and lemmatised single word frequency counts for nouns, non-lemmatised and lemmatised frequency counts for the collocations, item length (number of letters), and order of occurrence as fixed effects. In addition, I entered all possible second order interactions. To detect possible multi-collinearities among the main effects, the variance inflation factor (VIF) scores were calculated using the usdm package in R (Naimi, 2015). I had expected potentially high correlations between lemmatised and non-lemmatised frequency counts for nouns and collocations. The VIF-scores did not indicate any problem for non-lemmatised and lemmatised frequency counts for nouns (VIF = 2.28 and 1.91 respectively). However, they indicated a collinearity issue for non-lemmatised frequency counts for collocation and item type (VIF =51.91 and 54.31). I discarded the non-lemmatised frequency counts for collocations from the model. After fitting the maximal model, I started eliminating the predictor variables, one by one, that had the least impact on the model without making any distinctions between the main effects and second order interactions. I stopped the procedure when eliminating a predictor variable had not reduced the Akaike Information Criterion (AIC) value. After this, I visually inspected the residuals vs the fitted plot (see appendix H), which confirmed that the residuals were normally distributed.

According to the AIC values, the most parsimonious version of Model 2 included the item type, lemmatised collocation frequency, noun frequency, length, and nouns' order of occurrence as main effects. As can also be seen in Table 6.7, the identified model included item type by nouns' order of occurrence, and item length by noun frequency as interactions. The L1 Turkish participants responded to the high-frequency collocations faster than the (non-collocational) baseline items (β_1 = -.431, [SE= .044], p < .001). They also responded to the low-frequency collocations faster than the baseline items (β_2 = -.258, [SE= .038], p < .001). This finding confirms the results related to the item types in Model 1. The lemmatised collocation frequency counts significantly predicted log RTs (β_3 = .02, [SE= .009], p < .05). This indicates that as lemmatised collocation frequency counts increased, L1 Turkish participants needed more time to respond. Length of the items significantly predicted the log RTs (β_4 = .02, [SE= .005], p < .001). The order of occurrence of the nouns did not predict log RTs (β_5 =- .003, [SE= .012], p > 0.1), and neither did non-lemmatised noun frequency (β_6 = -.023, [SE= .037], p > .05). Furthermore, the identified model included the interactions between item type and nouns' order of occurrence; and item length by noun frequency. To make sure that these interactions contributed to the identified model, I ran a likelihood ratio test to compare two models, one with the main effects alone and the other with the main effects and the interactions together. The model comparison shows that there was a statistically significant difference between two models (chi-square = 9.75, p < .05).

Table 6.7 Selected mixed effect model 2 (language specific factors predicting RTs for Turkish L1 group)

Fixed effects	Estimate	Std. error	Df	*T*-value	*P*-value
Intercept	.268	.203	101	1.32	>0.1
High-frequency coll.	**-.431**	**.044**	**153**	**-9.78**	**<.001**
Low-frequency coll.	**-.258**	**.038**	**156**	**-6.77**	**<.001**
Lem. Coll. Freq.	**.020**	**.009**	**133**	**2.18**	**<.05**
Length	**.45**	**.20**	**99**	**2.22**	**<.05**
Order of Occurrence	-.003	.012	4010	-0.24	>0.8
Noun frequency	-.02	.038	98	-0.6	>0.5
Strong coll. * Order of Occ.	-.001	.02	4024	-0.05	>0.9
Weak coll. * Order of Occ.	**.047**	**.021**	**4028**	**2.19**	**<.05**
Length * Noun frequency	**-.079**	**.038**	**99**	**-2.03**	**<.05**

R^2 Marginal = .17, R^2 conditional = .42

To analyse the interaction between item type and nouns' order of occurrence, I conducted a post-hoc test using the least square means through the emmeans package in R (Lenth, 2018), with Tukey adjustments for multiple comparisons. The results showed that there was no statistically significant difference between RTs to the first vs the second occurrence of the nouns in the high-frequency collocations (β = .0044, [SE= .016], p = 0.9), in the low-frequency collocations (β = -.044, [SE= .017], p = 0.1), and or in the non-collocational (baseline) items (β = .0032, [SE= .012], p = 0.9). The interaction between item length and non-lemmatised noun frequency significantly predicted the log RTs (β = -.079, [SE= .038], p < .05). Using the visreg package (Breheny & Burchett, 2017), I visually inspected this interaction (item length by non-lemmatised noun frequency. It is possible to say that the interaction is weak. The item length seems to be interacting with the lower frequency nouns more strongly than the higher frequency nouns. Both the main effects (lemmatised noun frequency and adjective frequency) and the interactions (item type by lemmatised noun frequency; item type by adjective frequency; item type by lemmatised collocation frequency; and item type by noun frequency) were removed from Model 2 during the model selection process because of their low estimate values.

Model 3 was constructed to predict the log RTs in English by L1 speakers of English. One participant's data was lost because of a technical error. Therefore the final analysis was based on data from the remaining 30 participants from the L1 speakers of English. Similar to the previous two models (1 and 2), Model 3 had by-subject adjustments to the intercept as well as by-subject adjustments to item type. That is to say, I specified subject and item as random intercept as well as a by-subject random slope for item type. Model 3 included item types (high-frequency, low-frequency, baseline), single word frequency counts for adjectives, non-lemmatised and lemmatised single word frequency counts for nouns, non-lemmatised and lemmatised frequency counts for the collocations, (all frequency counts were log adjusted), item length (number of letters), and order of occurrence as fixed effects. In addition, I entered all possible second order interactions. Using the VIF scores, multicollinearities among the main effects were detected (Naimi, 2015). The VIF-scores did not indicate a collinearity issue for lemmatised and non-lemmatised frequency counts for nouns (VIF = 5.3 and 4.41, respectively). However, they indicated a collinearity problem for non-lemmatised frequency counts for collocations and item type (VIF =19.5 and 43.3, respectively). To solve this problem, I discarded non-lemmatised frequency counts for collocations from the model. Following the same procedure used in Model 2, I firstly fitted the maximal model. After this, I started eliminating the predictor variables, one by one, that had the least impact on the model without making any distinctions between the main effects and the second order interactions. I stopped the procedure when eliminating a predictor variable did not affect the AIC value. Finally, I visually inspected the residuals versus fitted plot (see appendix H), which confirms that the residuals were normally distributed.

According to the AIC values, the most parsimonious version of Model 3 included item type, lemmatised collocation frequency, lemmatised noun frequency, non-lemmatised noun frequency, adjective frequency, and item length as main effects. As can be seen in Table 6.8, the identified model included the following interactions: item type by lemmatised collocation frequency; item type by lemmatised noun frequency; item type by non-lemmatised noun frequency; item type by adjective frequency; and adjective frequency by item length. Unexpectedly, item types as a main effect did not predict log RTs. L1 English participants' responses to the high-frequency collocations (β_1 = .413 [SE= .422], $p > 0.1$), and to the low-frequency collocations were not significantly faster from the non-collocational items (β_2 = .879 [SE= .536], $p > 0.1$). Lemmatised collocation frequency did not predict log RTs (β_3 = .025 [SE= .022], $p > 0.1$), neither did lemmatised noun frequency counts (β_4 = .038 [SE= .074], $p > 0.1$). Moreover, non-lemmatised noun frequency (β_5 = -.060 [SE= .088], $p > 0.1$), and adjective frequency counts did not predict log RTs (β_6 = -.060 [SE= .088], $p > 0.1$). The length of the items significantly predicted log RTs (β_7 = .028 [SE= .007], $p < 001$). This indicated that items with more letters received longer log RTs. As can be seen in table 6.7, the following interactions: item type by lemmatised collocation frequency; item type by lemmatised noun frequency; item type by non-lemmatised noun frequency; item type by adjective frequency; and adjective frequency by item length all significantly predicted the log RTs. It should be noted that there is a statistically significant difference between main effect alone model and the model including both main effects and the interactions together (model comparison chi-square = 29.15, $p < .001$), according to the likelihood ratio test.

Table 6.8 Selected mixed effect Model 3 (language specific factors predicting RTs for L1 English participants)

Fixed effects	Estimate	Std. error	Df	*T*-value	*P*-value
Intercept	.281	.274	115	1.02	>0.3
High-frequency coll.	.413	.422	97	0.97	>0.3
Low-frequency coll.	.879	.536	118	1.64	>0.1
Lem. Coll. Freq.	.025	.022	124	1.13	>0.2
Lem Noun Freq.	.038	.074	137	0.51	>0.6
Noun Freq.	-.06	.088	130	-0.68	>0.4
Adj. Freq.	-.001	.016	110	-0.07	>0.9
Length	**.028**	**.007**	**103**	**3.71**	**<.001**
High-freq coll. * Lem. Coll. Freq.	**-.130**	**.063**	**90**	**-2.04**	**<.05**
Low-freq coll. * Lem. Coll. Freq.	**-.170**	**.083**	**111**	**-2.03**	**<.05**
High-freq coll. * Lem. Noun. Freq	-.035	.116	102	-0.3	>0.7
Low-freq coll. * Lem. Noun. Freq	**-.74**	**.023**	**128**	**-3.21**	**<.01**
High-freq coll. * Noun. Freq	-.022	0.12	104	-0.17	>0.8
Low-freq coll. * Noun. Freq	**.644**	**.234**	**127**	**2.75**	**<.01**
High-freq coll. * Adj. Freq	**.054**	**.024**	**97**	**2.22**	**<.05**
Low-freq coll. * Adj. Freq	.038	.023	116	1.65	>0.1
Adj Freq. *Length	**.016**	**.008**	**101**	**1.99**	**<0.05**

R^2 Marginal = .20, R^2 conditional = .38

To gain a clearer understanding of the results, and to more adequately address the hypotheses 3 and 4, I analysed the interactions as shown in Table 6.9, through the pairwise comparisons of item types. Table 6.9 shows the degree to which L1 English participants' log RTs were affected by the following variables: lemmatised collocation frequency; lemmatised noun frequency; non-lemmatised noun frequency; and adjective frequency counts for each item types (high-frequency, low-frequency and baseline). The interaction between the item types and the lemmatised collocation frequency showed that the high-frequency collocations (β_1 = -.130 [SE= .063], p <.05), and the low-frequency collocations were significantly different from the baseline items with respect to the degree to which lemmatised collocation frequency affected the log RTs (β_2 = -.170 [SE= .083], p <.05). This indicated that as the lemmatised collocation frequency counts increased in the high and low-frequency collocations, L1 English participants needed less time to respond. However, a re-levelling of the model from the baseline items to the strong collocations revealed that no significant difference was found in the comparisons of high and low-frequency collocational conditions between themselves (β_3 = -.040 [SE= .01], p >0.1).

The interaction between the item types and the lemmatised noun frequency showed that there was no significant difference between the high-frequency collocations (β_4 = -.035 [SE= .11], p > 0.1), and the baseline items with respect to the degree to which lemmatised noun frequency affected the log RTs. In contrast, a statistically significant difference was detected between the low-frequency collocations, (β_5 = -.741 [SE= .23], p < .01), and the baseline items and also between the low-frequency collocations, (β_6 = -.706 [SE= .23], p < .01), and the high-frequency collocations. That is to say, as lemmatised noun frequency increased in the low-frequency collocations, L1 English participants needed less time to respond. The interaction between the item types and the non-lemmatised noun frequency showed that the high-frequency collocations were not significantly different from the baseline items (β_7 = -.022 [SE= .12], p > 0.1), whereas the low-frequency collocations were significantly different from the baseline items - with respect to the degree to which non-lemmatised noun frequency affected the log

RTs (β_8 = .64 [SE= .23], p < .01). Re-levelling the model showed that the low-frequency collocations were also significantly different from the high-frequency collocations (β_9 = .66 [SE= .23], p < .01). This suggests that as the non-lemmatised noun frequency increased in the low-frequency collocations, L1 English participants needed more time to respond.

The interaction between the item types and the adjective frequency showed that the high-frequency collocations (β_{10} = .054 [SE= .024], p < .05) were significantly different from the baseline items - with respect to the degree to which adjective frequency affected the log RTs. No significant difference was found between the low-frequency collocations (β_{11} = .038 [SE= .023], p > 0.1), and the baseline items. Moreover, re-levelling the model showed no significant difference between the high-frequency collocations (β_{12} = -.016 [SE= .023], p > 0.1), and the low-frequency collocations. This suggests that as the adjective frequency increased in the high-frequency collocations, L1 English participants needed more time to respond. Table 6.9 also showed a significant interaction between adjective frequency and item length. Using the visreg package (Breheny & Burchett, 2017), I visually inspected this interaction (adjective frequency by length. It should be noted that the interaction between adjective frequency and length significantly predicted the log RTs (β_{13} = .016 [SE= .008], p >.05). The interaction is overall weak. Both the main effect (noun's order of occurrence) and the interactions (item type by length; lemmatised collocation frequency by length; and item type by noun's order of occurrence, lemmatised noun frequency by item length and non-lemmatised noun frequency by item length) were removed from the model during the model selection process because of their low estimate values.

Table 6.9 Multiple comparison table for item types and frequency counts

Contrasts	Predictor	Estimate	Std. error	Df	*T*-value	*P*-value
High freq-Baseline	Lem. Coll. Freq.	**-.130**	**.063**	**90**	**-2.04**	**<.05**
Low-freq-Baseline	Lem. Coll. Freq.	**-.170**	**.083**	**111**	**-2.03**	**<.05**
Low freq-High-freq	Lem. Coll. Freq.	-.040	.102	101	-0.39	>0.6
High freq-Baseline	Lem. Noun. Freq.	-.035	.116	102	-0.3	>0.7
Low freq-Baseline	Lem. Noun. Freq.	**-.741**	**.230**	**128**	**-3.21**	**<.01**
Low freq-High freq	Lem. Noun. Freq.	**-.706**	**.236**	**119**	**-2.98**	**<.01**
High freq-Baseline	Noun Freq.	-.022	.125	104	-0.17	>0.8
Low freq-Baseline	Noun Freq.	**.644**	**.234**	**127**	**2.75**	**<.01**
Low freq-High-freq	Noun Freq.	**.667**	**.236**	**120**	**2.82**	**<.01**
High freq-Baseline	Adjective Freq.	**.054**	**.024**	**97**	**2.22**	**<0.5**
Low freq-Baseline	Adjective Freq	.038	.023	116	1.65	>0.1
Low freq-High freq	Adjective Freq	-.016	023	101	-0.69	>0.4

As predicted, L1 Turkish-English L2 (advanced) participants RTs' were faster for the high-frequency collocations (high-frequency: 943 ms (SD =383)) compared to the low-frequency collocations (low-frequency: 1146 ms (SD =477)), and to the baseline items (baseline = 1326 ms (SD =559)). The same pattern was also seen in English L1 participants' RTs. They were faster for the high-frequency collocations (high-frequency: 892 ms (SD =338)) compared to the low-frequency collocations (low-frequency: 1075 ms (SD =431)), and to the baseline items (baseline = 1303 (SD =527)). Model 4 was constructed to predict the log RTs by L1 speakers of English and L1 Turkish-English L2 participants in English. Similar to the previous models 1, 2, and 3, Model 4 had by-subject adjustments to the intercepts as well as by-subject adjustments to item type. Specifically, the model had subject and item as random intercepts as well as a by-subject random slopes for item type. Model 4 included group (L1 English, L1 Turkish-English L2), item type (high-frequency, low-frequency, baseline), single word frequency counts for adjectives, non-lemmatised and lemmatised single word frequency counts for nouns, non-lemmatised and lemmatised frequency counts for the collocations, (all frequency counts were log adjusted), item length (number of letters), and order of occurrence as fixed effects. I also entered all possible second order interactions. Using the VIF scores, multicollinearities among the main effects were detected (Naimi, 2015). The VIF-scores did not indicate a collinearity issue for lemmatised and non-lemmatised frequency counts for nouns (VIF = 4.21 and 4.30, respectively). However, they indicated a clear collinearity problem for non-lemmatised and lemmatised frequency counts for collocation and item type (VIF =20.16, 14.17 and 35.01, respectively). To solve this problem, I discarded both non-lemmatised and lemmatised frequency counts for collocations from the model. Following the same procedure used in Models 2 and 3, I firstly fitted the maximal model. After this, I started eliminating the predictor variables, one by one, that had the least impact on the model without making any distinctions between the main effects and the interactions. I stopped the procedure when eliminating a predictor variable did not decrease the AIC value. Finally, I visually inspected the residuals vs fitted plot (see appendix H), which confirms that the residuals were normally distributed.

According to AIC values, the most parsimonious version of model 4 included, group, item type, noun frequency, and item length as main effects (see Table 6.10). The identified model included the following interactions: group by item type; group by noun frequency; and group by length. The L1 English participants did not appear to respond to the collocations and non-collocational items significantly faster than L1 Turkish-English L2 (advanced) participants (β_1

= .185, [SE= .165], $p > 0.1$). Both L1 English and L1 Turkish-English L2 participants responded to the high-frequency collocations faster than the (non-collocational) baseline items (β_2 = -.363, [SE= .024], $p < .001$). They also responded to the low-frequency collocations faster than the non-collocational items (β_3 = -.180, [SE= .023], $p < .001$). Re-levelling the model from baseline items to the low-frequency collocations showed that both groups of participants responded to the high-frequency collocations faster than the low-frequency collocations (β= -.183, [SE= .021], $p < .001$). This finding was further investigated by analysing the interaction between group and item type (see Table 6.11). Noun frequency significantly predicted the log RTs (β_4 = -.072, [SE= .028], $p < .05$). This indicated that as the noun frequency increased, participants needed less time to respond. Unsurprisingly, the length of the items significantly predicted the log RTs (β_5 = .033, [SE= .007], $p < .001$). The likelihood ratio test showed that there was no statistically significant difference between the main effect only model and the model including both main effects and the interactions together (model comparison chi-square = 5.44, $p > 0.1$).

To more adequately address the hypothesis 5 about L1 English and L1 Turkish-English L2 (advanced) participants' sensitivity to collocational frequency and single-word frequency scores, I carried out a post-hoc test using the least square means through the emmeans package in R (Lenth, 2018), with Tukey adjustments for multiple comparisons. As can be seen in Table 6.11, the results showed that L1 English participants responded to the low-frequency collocations faster than the baseline items (β_1 =-.192, [SE= .029], $p < .0001$), and they responded to the high-frequency collocations faster than the low-frequency collocations (β_3 =-.183, [SE= .024], $p < .0001$). L1 Turkish-English L2 participants' responses followed the same pattern. They responded to the low-frequency collocations faster than the baseline items (β_4 =-.167, [SE= .028], $p < .0001$), and they responded to the high-frequency collocations faster than the low-frequency collocations (β_6=-.183, [SE= .024], p < .0001). In this post-hoc test, I also compared L1 English and L1 Turkish-English L2 (advanced) participants' RTs for the two collocational conditions and the baseline items to find out if L1 speakers' RTs were significantly faster for any of these item types. The results indicated that L1 English participants' responses to the high-frequency collocations were not significantly faster than L1 Turkish-English L2 participants' responses to the high-frequency collocations (β=-.046, [SE= .038], p > 0.1). Furthermore, L1 English participants' responses to the low-frequency collocations were not significantly faster than L1 Turkish-English L2 participants' responses to the low-frequency collocations (β=-.046, [SE= .041], p > 0.1), neither their responses to the baseline items were significantly faster than the L1 Turkish-English L2 participants' responses to the baseline items (β=-.021, [SE= .053], p > 0.1).

Table 6.10 Mixed effect model 4 (Comparing L1 English and L2 English groups for item types)

Fixed effects	Estimate	Std. error	Df	*T*-value	*P*-value
Intercept	**.616**	**.155**	**124**	**3.96**	**<.001**
L1 English	.185	.165	3178	1.11	>0.2
High-frequency coll.	**-.363**	**.024**	**117**	**-15.04**	**<.001**
Low-fequency coll.	**-.180**	**.023**	**117**	**-7.76**	**<.001**
Noun Freq.	**-.072**	**.028**	**118**	**-2.55**	**<.05**
Length	**.033**	**.007**	**117**	**4.59**	**<.001**
Group*Itemtype-High-freq	.024	.037	62	0.66	>0.5
Group*Itemtype-Low-freq	.024	.034	61	0.72	>0.4
Group*Noun Freq.	-.030	.029	5784	-1.04	>0.2
Group*Length	.013	.007	5777	1.84	>.05

R^2 Marginal = .18, R^2 conditional = .42

Table 6.11 Post-hoc test for within group comparison (item types)

Contrasts	Group	Estimate	Std. error	z. ratio	*P*-value
Low freq-Baseline	NS	-.192	.029	-6.61	<.0001
High freq-Baseline	NS	-.376	.030	-12.23	<.0001
High freq-Low freq	NS	-.183	.024	-7.38	<.0001
Low freq-Baseline	NNS	-.167	.028	-5.86	<.0001
High freq-Baseline	NNS	-.351	.030	-11.65	<.0001
High freq-Low freq	NNS	-.183	.024	-7.47	<.0001

6.4 Discussion

This study firstly explored the processing of adjective-noun collocations in Turkish and English by L1 speakers of these languages. The results showed that both L1 groups processed the high-frequency collocations faster than the low-frequency collocations, and the low-frequency collocations faster than the baseline items. Furthermore, adjective-noun collocations (which have a similar collocational frequency and similar association scores) were processed at comparable speeds by L1 speakers of English and Turkish. Nevertheless, L1 speakers of English and Turkish appeared to attend to noun frequency counts at different levels. This study secondly investigated the processing of adjective-noun collocations in English by L1 speakers of English and L1 Turkish-English L2 (advanced) speakers. The results showed that both L1 English, and L1 Turkish-English L2 advanced speakers processed the high-frequency collocations faster than the low-frequency collocations, and the low-frequency collocations faster than the baseline items. Both groups were sensitive to the single-word frequency counts of nouns and collocation frequency counts. The following sections 6.3.1 and 6.3.2 discuss these findings in detail.

6.4.1 Processing collocations in L1.

At the outset of this study, 5 research hypotheses were proposed. Hypothesis 1 stated that L1 speakers of both Turkish and English, would process the high-frequency collocations faster than the low-frequency collocations. In turn, they would also process the low-frequency collocations faster than the (non-collocational) baseline items. The results of Model 1 showed that both L1 groups processed the high-frequency collocations faster than the low-frequency collocations, and they also processed the low-frequency collocations faster than the baseline items in their respective native languages (see Table 6.5, and Table 6.6). This indicates that L1 speakers of both English and Turkish need less time to process the collocations which occur more frequently, and which are more strongly associated, than to the collocations which occur less frequently, and which are less strongly associated. Thus, collocational frequency and association are strong predictors for processing collocations in different languages. Similar results have been reported in a number of studies using participants in different first languages, and different methodologies. Durrant and Doherty (2010), using L1 English participants, found priming effects in the processing of both frequent and associated frequent collocations in comparison to non-collocations, employing a primed LDT task. In examining a different kind of MWS, (compositional phrases) Arnon and Snider (2010), using a phrasal decision task, reported that high-frequency phrases were processed faster than low-frequency phrases by L1 English adult speakers. Turning to Turkish collocations, Cangir et al. (2017), using a primed LDT, found that Turkish adjective-noun, and verb-noun collocations were responded to faster than the items in non-collocational conditions by adult L1 speakers of Turkish. Exploring an alternative methodology, McDonald and Shillcock (2003) and Vilkaite (2016), looking at L1 English participants' eye-movements, reported that they read verb-noun collocations faster control pairs. Additionally, Siyanova-Chanturia et al. (2017) using the ERP method, found that binomials were processed faster than novel phrases by L1 English speakers. Overall, the current study adds to the growing body of evidence that MWS are processed more quickly than novel phrases.

Hypothesis 2 stated that L1 speakers of Turkish will process both high-frequency and low-frequency collocations faster than L1 speakers of English. The results of mixed effect Model 1 also indicated that no significant differences were found in the participants' responses to the high-frequency and low-frequency collocations when the RTs in English and Turkish were compared. That is to say, high-frequency collocations were responded to equally quickly in English and in Turkish by L1 speakers of each language. Similarly, low-frequency collocations were responded to equally quickly in English and in Turkish. To the best of my knowledge, this is the first empirical evidence showing that adjective-noun collocations (which have a similar collocational frequency and similar association scores) are processed at comparable speeds by L1 speakers of different languages. This finding suggests that L1 users of typologically different languages are equally sensitive to collocational frequency counts. This is crucial in supporting the elevated status of MWS (specifically collocations) as a general feature of language, rather than merely a feature of a few select languages. It should be noted that this finding is in line with usage-based models of language (Ellis, 2002; Christiansen & Chater, 2016; Kemmer & Barlow, 2000; Tomasello, 2003), according to which frequency at across multiple grain sizes plays a key role in language acquisition and processing. Furthermore, this finding, gathered from on-line processing of collocations in typologically different languages, adds empirical support to the growing body of research indicating that language users rely on MWS in language comprehension and production (e.g. Arnon and Snider 2010; Arnon et al. 2017; Bannard & Matthews, 2008; Siyanova-Chanturia et al. 2017).

Looking at the effect of non-lemmatised collocation frequency counts on the processing of MWS has been the common practice for psycholinguistic studies (Vilkaite, 2016, Wolter & Gyllstad, 2011; 2013; Wolter & Yamashita; 2015, 2018). Nonetheless, considering the typological differences between Turkish and English, looking at the effects of both non-lemmatised and lemmatised collocation frequency scores on the participants' RTs provided a more comprehensive understanding of how language typology affects the processing of collocations in these languages. In this regard, hypothesis 3 stated that lemmatised collocation-level frequency counts would have a larger effect on the processing of Turkish collocations than English collocations. The results of mixed effect Model 2 showed that as the lemmatised collocation frequency counts increased, L1 Turkish participants needed more time to respond to Turkish collocations (see Table 6.7). In contrast, mixed effect Model 3 indicated that, as the lemmatised collocation frequency counts increased, L1 English participants needed less time to respond to collocations (see Table 6.9). In other words, lemmatised collocation frequency counts predicted the RTs for English collocations in the same direction as non-lemmatised collocation frequency counts. However, lemmatised collocation frequency counts predicted the RTs for Turkish collocations in the opposite direction as non-lemmatised frequency counts. This result is in line with the findings from the corpus study discussed in Section 4.3.1. The corpus study has provided evidence that Turkish collocations have considerably larger differences between their lemmatised and unlemmatised collocational frequency scores than the English collocations. When the lemmatised noun and lemmatised collocation frequency scores are used to calculate the MI and LD scores, some of the Turkish adjective-noun collocations had considerably lower MI and LD scores than the *non*-lemmatised collocations (see Table 4.5, and Table 4.6, see section 4.3.1 for a discussion individual examples). The possible reason for the lower association scores for lemmatised collocations was that not all of the inflected forms of the noun collocate with their adjectives to the same degree as the base forms. In other words, when the Turkish adjective-noun collocations are lemmatised, both the noun and collocation frequency increase with the addition of the inflected forms, but if the inflected forms do not collocate with the adjective to a similar degree as the base form, their collocational strength becomes weaker than the non-lemmatised forms.

It would be interesting to compare the differing effects of lemmatised and non-lemmatised collocation frequency counts on the processing of collocations found in this study with other psycholinguistic experiments which look at the processing of MWS in other pairs of languages. However, to the best of my knowledge, no psycholinguistic work focusing on the processing of MWS has explored the effect of lemmatised frequency counts. The only comparable study is Durrant's (2014), which examined the effects of both lemmatised and non-lemmatised collocation frequency counts on L2 English learners' collocational knowledge in a meta-analysis. He reported that there was "no clear differences" between the lemmatised and non-lemmatised counts of collocation frequency scores for predicting learners' collocational knowledge. Furthermore, he added that the differences in span (4:4 vs 9:9) did not affect the results. It should be noted that Durrant's (2014) finding related to lemmatised collocation frequency counts' predicting collocational knowledge is in line with the finding in this study on L1 English speakers' needing less time to respond as the lemmatised collocation frequency counts increased. In this regard, Turkish L1 speakers' needing more time to respond as the lemmatised collocation frequency counts increase can be attributed to the agglutinating structure of Turkish. Clearly, this finding should be seen as tentative, and to be explored further through psycholinguistic experiments comparing the processing of MWS in agglutinating and non-agglutinating languages.

Hypothesis 4 stated that word-level frequency counts of the nouns will have a larger effect on the processing of Turkish adjective-noun collocations than English adjective-noun collocations. It was surprising that L1 Turkish speakers attend to noun frequency counts to a notably lesser extent than L1 English speakers. The reason it was hypothesised that L1 Turkish speakers would be more sensitive to noun frequency counts was that individual word forms in Turkish have considerably lower frequency counts than similar word forms in English (Durrant, 2013). The likely reason for this lower frequency of individual word forms in Turkish is that meanings which are expressed using multi-words in English, can be conveyed using various inflected forms of single-word forms in Turkish. Therefore, it was predicted that L1 Turkish speakers would be more sensitive to the noun frequency counts than English L1 speakers. Surprisingly single word frequency counts for nouns did not predict L1 Turkish speakers' RTs for either high-frequency or low-frequency collocations. However, they appear to affect the L1 English speakers' processing of low-frequency adjective-noun collocations significantly. In other words, as lemmatised noun frequency increased in the low-frequency collocations, L1 English participants needed less time to respond. Notably, noun frequency counts do not appear to significantly affect the L1 English speakers' RTs for high-frequency collocations. A possible reason for this difference is that L1 English speakers appear to attend to collocation frequency rather than single word frequencies as the collocation frequency of the items increase (see Table 6.9).

In addition to noun frequency, adjective frequency counts also affected L1 English speakers' processing of adjective-noun collocations. As the adjective frequency increased in the high-frequency collocations, L1 English participants needed more time to respond. This is an expected finding from a collocational processing perspective, because when participants see a collocation which includes a very frequent adjective (e.g *long time*), predicting the upcoming noun, the next word, would be more difficult for the L1 English speakers. Similar results about L1 English speakers' being sensitive to both collocational and single-word frequency counts have also been reported previously. Wolter and Yamashita (2018) found that L1 English speakers relied more heavily on collocation frequency counts than word-level frequency, but they were still sensitive to the word-level frequency counts of collocations. Adding to the finding of Wolter and Yamashita (2018) about L1 speakers' attending to collocation and single-word frequency counts simultaneously, this study provided evidence that frequencies of collocations influence the degree to which speakers attend to single-word and collocation frequency counts. Another important contribution of the present study is that the highly frequent individual words particularly adjectives within collocations might also negatively affect the processing speed of collocations since they are likely lead to weakly associated collocations (see Table 6.9).

The finding that L1 Turkish speakers' processing of collocations is not significantly affected by noun and adjective frequency counts provides empirical evidence that L1 Turkish speakers tend to process collocations more holistically than L1 English speakers. Wray (2008, p.12) put forward the definition of "morpheme equivalent unit", a string of words processed like a morpheme, that is, without meaning-matching of any sub-parts. In line with this view, some studies have reported evidence for holistic processing of MWS (e,g. Jiang & Nekrasova, 2007; Underwood, Schmitt, & Galpin, 2004). As it was previously noted, holistic storage and processing implies that MWS are processed as a holistic unit without access to or analysis into its constituents (Siyanova-Chanturia, 2015), (see sections 1.1, 1.2, 1.3 for more detailed review). It should be noted that the current study has provided evidence that both adjectives and nouns as individual parts of the collocations affected the processing of collocations for L1 English speakers (see Table 6.9). Looking at these findings related to the effects of single-word frequency counts on collocational processing, Turkish L1 speakers' processing of Turkish adjective-noun collocations seem to be more holistic than English L1 speakers' processing of English adjective-noun collocations. The corpus-based study (see Chapter 4) showed that a processing model in which adjective-noun collocations are either holistically stored or fully processed does not seem to be well-suited to accounting for the processing of collocations in Turkish. The reason is that although the vast majority of the inflected forms of the collocations have lower collocational frequency counts and LD-scores than their base forms, some of the inflected forms have quite high LD and collocational frequency scores (e.g. *kısa zamanda*, with LD-score of 8.97). Therefore, they are also likely to be enjoying processing facilitation. In addition, as also argued by Durrant (2013), any processing model in which grammar and lexis are treated as fully independent systems (Pinker, 1999), seem to be poorly suited to account for these data. The usage-based approaches to language acquisition (Christiansen & Chater, 2016a; Ellis, 2002; Tomasello, 2003) is more consistent with the psycholinguistic and corpus data of formulaicity of adjective-noun collocations, according to which language users are sensitive to the frequency information at multiple grain sizes.

To recap briefly, this study showed that collocations which have similar collocational frequency and association scores are processed at comparable speeds in English and Turkish by L1 speakers of these languages. Furthermore, lemmatised collocation frequency counts affected the processing of Turkish and English collocations differently, and Turkish speakers appeared to attend to word-level frequency counts of collocations to a lesser extent than English speakers. In the following section, the findings of the study related to L1 and L2 processing of collocations are discussed.

6.4.2 Processing collocations in the L2.

In addition to L1 English and L1 Turkish processing of collocations in their native-languages, this study also examined whether there would be differences in L1 English and L1 Turkish English L2 advanced speakers' processing of collocations in English, with respect to attending to collocational and single-word frequency counts. Hypothesis 5 stated that L1 Turkish-English L2 speakers (advanced) will show sensitivity to both collocational and single-word level frequency counts. The results of mixed effect Models 4 indicated that L1 English speakers did not appear to process the high-frequency and low-frequency collocations significantly faster than L1 Turkish-English L2 (advanced) speakers (see Table). This finding was also confirmed by the results of the post-hoc test that there was not any significant interaction between group and item type. Furthermore, the results of mixed effect Models 4 also revealed that both groups (L1 English and L1 Turkish-English L2) speakers processed the high-frequency collocations faster than the (non-collocational) baseline items, and low-frequency collocations faster than the baseline items in English (see Table 6.10 and Table 6.11). These findings are in line with hypotheses 5 that L1 Turkish-English L2 advanced speakers are sensitive to the collocation frequency counts like L1 speakers of English because as the collocation frequency counts increased, L1 Turkish-English L2 speakers responded to the collocations in English more quickly, as L1 English speakers did. The results of the current study add to the growing body of research showing that advanced level L2 speakers are sensitive to the frequency distribution of frequently occurring MWS, just as L1 speakers are (Siyanova-Chanturia et al, 2011; Wolter & Gylstad, 2013; Wolter & Yamashita, 2017). It is noteworthy that only the congruent collocations were preferred in item development stage, considering the empirical evidence that L1 congruency affects the processing of collocations in L2 (Wolter & Gylstad, 2013; Wolter & Yamashita, 2017).

This study also explored L1 English and L1 Turkish English L2 advanced speakers' attending to single-word frequency counts alongside collocation frequency counts. The results of mixed effect Models 4 indicated that both groups (L1 English and L1 Turkish-English L2 advanced speakers) attended to noun frequency counts, and there was no significant interaction between group and noun frequency counts. This indicates that noun frequency counts predicted the RTs for both groups, and L1 Turkish English L2 advanced speakers did not appear to rely on the noun frequency scores more heavily than the L1 English group while reading adjective-noun collocations in English. This finding is in line with the hypothesis 5 that L1 Turkish-English L2 advanced speakers are sensitive to the noun frequency counts like L1 speakers of English. As the noun frequency increased, both groups needed less time to respond. Similar results have been reported by Wolter and Gyllstad, (2018). They found that both L1 and L2 speakers of English attend to word-level frequency and collocational frequency counts simultaneously. The difference between the findings of Wolter and Yamashita, (2018), and the current study is that Wolter and Yamashita (2018) found that a difference between L1 speakers and intermediate level L2 speakers relying on the collocation level versus single-word level frequency counts. They reported that L1 English and L2 English advanced groups appeared to attend to collocation frequency more than single-word level frequency counts. However, intermediate level L2 speakers seemed to attend to single-word frequency counts more than collocational frequency counts. They also found that advanced L2 group attended to single-word frequency counts more than L1 English group. The current study however, is not in a position to observe the degree to which L1 and L2 speakers of different proficiency levels relying on single-word level and collocation level frequency counts since this study recruited only one group of L2 speakers, at a fairly advanced level speakers.

These findings (both L1 and L2 speakers' attending to both single-word level and collocation level frequency) counts simultaneously is in contrast to Wray's (2002, 2008) claims that L1 and L2 speakers process the MWS in fundamentally different ways (see section 1.3 for a detailed discussion on Wray's approach). The results of both the current study and some other psycholinguistic experiments (e.g. Siyanova-Chanturia et al. 2011; Wolter and Gyllstad, 2018) suggest a more unified approach in which all language users (including L2 speakers with a certain level of proficiency) are sensitive to the frequency information at different grain sizes (both single-word, and multi-word levels). More generally, a speaker's life-time experiences of language appear to affect the processing of collocations in both L1, and L2. In other words, these findings are in line with usage-based models of language (Ellis, 2002; Christiansen & Chater, 2016; Kemmer & Barlow, 2000; Tomasello, 2003), according to which frequency, and probability of input plays a key role in language acquisition and processing (see Section 1.5 and for a detailed discussion on usage-based models of language).

6.4.3 Limitations.

One important limitation of the current study is that the AJT task used in the experiment did not make it possible to compare the potential differences of priming effects between adjective-noun collocations in Turkish and English. Since the effects of noun frequency counts appear to be smaller for processing Turkish adjective-noun collocations, observing the differences of priming effects between adjectives and nouns in two languages, would potentially shed more light on the processing of collocations in agglutinating languages. Another important limitation of the present study is that it did not include inflected adjective-noun combinations alongside the base forms in the experiment, so that it failed to address an important question that whether the inflected forms of the collocations also enjoy processing facilitation (see also section 7.5 for a further discussion on the limitations of the current study). It would also be helpful to do a norming study for the semantic transparency of the items used in the experiments before carrying out the actual study. Since I have not recruited tranined native-speakers of English and Turkish to rate the semantic transparency of collocations used in the AJTs, it is not easy to say at this point whether semantic transparency of the collocations used significantly affected the processing speeds. Furthermore, it would provide further insight to analyse the collocations and baseline items separately since the baseline items required a no response whereas the collocations both (high-frequency and low-frequency) required a yes response.

6.5 Conclusion

In conclusion this part of the thesis firstly provided a detailed literature review of L1, and L2 processing of MWS. Then, it reported on a psycholinguistic experiment, which sets out to investigate the processing of adjective-noun collocations in Turkish and English. More specifically, it examined the influence of both single word, and collocational frequency counts on the processing of collocations in Turkish and English by L1 speakers of these languages. Furthermore, this study explored the variables which affects the processing of adjective-noun collocations by L1 Turkish-English L2 advanced speakers in English. The AJTs were used to investigate the processing of collocations in Turkish by L1 speakers and English by both L1 English and L1 Turkish L2 English speakers. The results showed that that adjective-noun collocations (which have a similar collocational frequency and similar association scores) are processed at comparable speeds by L1 speakers of different languages. This finding suggests that language users of typologically different languages are equally sensitive to collocational frequency counts. Furthermore, lemmatised collocation frequency counts affected the processing of Turkish and English collocations differently, and Turkish speakers appeared to attend to word-level frequency counts of collocations to a lesser extent than English speakers.

These differences highlight that L1 speakers of English and Turkish do not process the adjective-noun collocations identically. Furthermore, both L1 English, and L1 Turkish-English L2 advanced speakers are sensitive to the collocation frequency counts and noun frequency simultaneously. This indicated that both groups are sensitive to the frequency information at different grain sizes, and this finding is compatible with usage-based model of language, according to which, frequency of input plays a key role in language acquisition, and processing.

Notes

1- I excluded the studies focusing on the processing of idiomatic expressions (e.g. *kick the bucket*) from the literature review in chapter 1. The theories used to account to for idiomatic phrases might not necessarily shed light on the processing of other type of MWS (e.g. collocations)

2- I intended to systematically introduce random slopes in the models to include item based variations. However, introducing random slopes for item types led to the convergence errors. Following the suggestions in Baayen et al. (2008), I simplified the models.

Part 4: General Discussion and Conclusion

This thesis so far provided theoretical and methodological frameworks (see chapters 1 and 2). Furthermore, it reviewed the previous literature of empirical studies focussing on MWS from a corpus lingiustics and psycholinguistic perspectives (see chapter 3 for a review of corpus studies and chapter 5 for a review of psycholinguistic experiments). Chapter 4 reported on a corpus study examining the frequency counts and association statistics of adjective-noun collocations in English and Turkish from a contrastive perspective. Chapter 6 reported on psycholinguistic experiments investigating the processing of adjective-noun collocations in English and Turkish by L1 speakers of these languages, and the processing of adjective-noun collocations in English by L1 English and L1 Turkish-English L2 advanced level speakers (see section 2.3.3 for research questions of these studies). In this concluding part, chapter 7 discusses the findings of the corpus, and experimental studies. Furthermore, it discusses the methodological approach of this thesis, focussing on combining the corpus and psycholinguistic experimentation. Then, it summarises the limitations of the present thesis and provides detailed directions for future research on the processing and using MWS by L1 and L2 speakers. Finally, it presents the concluding remarks for this thesis.

Chapter 7: Interpration of the findings

The results reported in this thesis revealed that the present study has made both theoretical and methodological contributions to the fields of psycholinguistics with regard to L1 and L2 processing of MWS (see sections 4.2 for the findings of the corpus study and 6.2 for the findings of the experimental study). Therefore, this chapter firstly attempts to interpret the corpus and experimental findings together in relation to L1 processing of adjective-noun collocations in agglutinating and non-agglutinating languages (see section 7.1). Secondly, this chapter discusses the L1 and L2 processing of collocations (see section 7.2). Thirdly, it discusses the methodological implications of this thesis. More specifically, it presents how this study combined corpus and experimental psycholinguistics methods to study the processing of collocations (see section 7.3). Finally, it presents the limitations of the current study, and presents some detailed future directions (see section 7.4).

7.1 L1 processing of collocations in typologically different languages: Collating the findings of the corpus study and psycholinguistic experiments

Collocations and other type of MWS are predicted to have, (on average), lower frequency scores in Turkish than in English. However, the corpus study suggested that unlemmatised collocations in both frequency bands have similar mean frequency and association counts as measured by MI and LD measures in the both languages (see section 4.3.1). This suggests that the base forms of the collocations in English and Turkish do not appear to have notably different frequency and association counts from each other. To test the effect of agglutinating structure of Turkish on the collocability of adjectives and nouns, the lemmatised forms of the collocations in the both languages were examined. The findings indicated that the vast majority of the lemmatised Turkish adjective-noun combinations occur at a higher-frequency than their English equivalents. In addition, lemmatised collocations' LD-scores showed that agglutinating structure of Turkish appears to increase adjective-noun collocations' association scores in the both frequency bands since the vast majority of Turkish collocations reach higher scores of collocational strengths than their unlemmatised forms. Nevertheless, it would be mistaken to assume that agglutinating structure of Turkish increases the collocational strength of all Turkish collocations. The inflected form of some of these collocations have considerably lower association scores than their base forms (see 4.3.1 for a detailed discussion on these individual examples of collocations).

Having seen the corpus evidence that agglutinating structure of Turkish appears to make a positive contribution to the collocational strength of the vast majority of the adjective-noun collocations examined, the AJTs also investigated the similarities and differences between L1 English and Turkish speakers' on-line processing of adjective-noun collocations in their respective first languages (see chapter 6). The results showed that adjective-noun collocations were responded to equally quickly in English and in Turkish by L1 speakers of each language. To the best of my knowledge, this is the first empirical evidence indicating that adjective-noun collocations (which have a similar collocational frequency and strength) are processed at comparable speeds by L1 speakers of typologically different languages. The findings suggest that speakers of typologically different languages are equally or at least similarly sensitive to collocational frequency counts. From a corpus linguistics perspective, this is not a totally unexpected finding since the base (unlemmatised) forms of the collocations in English and Turkish seem to have quite similar frequency and association counts. Furthermore, one important theoretical implication of this finding gathered from on-line processing of collocations in typologically different languages is that it adds empirical support to the growing body of evidence that usage-based approaches to language acquisition (Ellis, 2002; Christiansen & Chater, 2016b; Kemmer & Barlow, 2000; McCauley & Christiansen, 2016a), which view linguistic productivity as a gradually emerging process of storing and abstracting MWS.

Although, L1 English and Turkish processing of collocations look similar with regard to their reliance on collocational frequency counts (see Table 6.5, and Table 6.*6*), the findings of the experiments revealed some interesting differences with regard to their attending to single-word frequency counts of adjectives and nouns. Noun frequency counts did not significantly predict L1 Turkish speakers' RTs. However, lemmatised and non-lemmatised noun frequency counts significantly interacted with the item type for English L1 participants' RTs. More specifically, as noun frequency counts increased in the low-frequency collocations, L1 English participants needed less time to respond. Importantly, noun frequency counts do not appear to significantly affect the L1 English speakers' RTs for high-frequency collocations. In addition to noun frequency counts, L1 English speakers showed sensitivity to the adjective frequency counts. As the adjective frequency increased in the high-frequency collocations, L1 English participants needed more time to respond. This is an expected finding from a corpus linguistics perspective, because frequent adjectives tend to lead to weakly associated collocations (e.g *long time*), rather than strongly associated ones. Furthermore, from a collocational processing perspective when participants see a very frequent adjective, predicting the upcoming noun, the next word, would be more difficult for the L1 English speakers. However, L1 speakers of Turkish does not seem to show sensitivity to frequency counts of adjectives while processing adjective-noun collocations. Based on these findings, it is possible to say there is a clear difference between L1 English and Turkish speakers' attending to single word frequency counts for processing adjective-noun collocations.

The variation in L1 English and Turkish speakers' attending to single word frequency counts of adjectives and nouns is an important finding since it has a potential to contribute to the wider discussion about whether MWS are processed holistically that is without access to their individual words (Wray, 2002; 2008, see also sections 1.1, 1.2, and 1.3 for a detailed discussion on holistic processing and storage of MWS). The findings of the current study provided evidence that speakers are more sensitive to the frequent collocations. In other words, collocations gradually gain greater prominence as a chunk in relation to the individual components. However, single word frequency counts for nouns predict L1 English speakers' RTs for low-frequency collocations. Therefore, the individual components of the MWS appear to be still accessible for L1 English speakers. The current study is not alone in this view of L1 English speaker's attending to single word and collocational or phrasal frequency counts at the same time. For example, Wolter and Yamashita (2018) found that L1 English speakers attend to both collocation frequency and word-level frequency counts, but they more heavily relied on the collocational frequency counts. Furthermore, Arnon and Cohen Priva (2015) found a change in the prominence of single word and multi-word information with growing phrase frequency. That is, for high-frequency trigrams, the effect of individual word frequency on phonetic duration decreased but it critically remained significant. However, the effect of phrase frequency on phonetic duration increased. Therefore, the results of Arnon and Cohen Priva (2015) study on the production of MWS is in line with the current study that frequent usage leads to a growing prominence of MWS; however, frequent usage of MWS does not fully eliminate the effect of single word frequency within MWS.

Interestingly, L1 Turkish speakers did not seem to attend to noun frequency counts either for high-frequency or for low-frequency collocations. Based on this finding, it is possible to say that L1 Turkish speakers tend to process adjective-noun collocations more holistically than L1

English speakers do in their L1. Nevertheless, it is important to bear in mind that the current experimental study was not designed to test the view of MWS being processed as unanalysed wholes. Since the design of the current experiment does not involve a collocational, semantic or syntactic priming paradigm or specifically look at the effect of individual words on the processing of MWS employing an eye-movement paradigm (see also Siyanova-Chanturia, 2015), the findings should be seen cautiously for interpreting them beyond the speed of processing. Therefore, for an alternative explanation, looking at the word processing literature in agglutinating languages might be helpful. Although these studies work within a dualistic paradigm, in which inflected words are either stored as full-form or accessed through their constituents, which is not compatible with usage-based models of language, they provide some important insights about lexical processing in agglutinating languages. Gürel (1999), using a simple LDT, tested the word recognition of morphologically simple and complex nouns in Turkish. She found that there was no significant difference between L1 Turkish speakers' RTs for monomorphemic and multimorphemic words with frequent suffixes. She interpreted this finding in such a way that L1 Turkish speakers are sensitive to the frequency of suffixes. Lehtonen and Mani (2003) also used an LDT task to examine the effects of frequency on morphological processing in Finnish nouns. They found no significant difference between L1 Finnish speakers' RTs for the high-frequency monomorphemic and multi-morphemic nouns. These two studies provided empirical evidence that speakers of agglutinating languages process the nouns differently from the speakers of non-agglutinating languages to be able to effectively handle the complexity of the word forms in these languages. In this regard, it is possible to predict that since speakers of agglutinating languages do not only attend to single words and MWS, but also to the morphemes attached to the words, they might rely heavily on the priming relationship between nodes and collocates within high-frequency collocations.

The current study also identified some notable differences between L1 English and Turkish speakers' attending to lemmatised collocational frequency counts while processing adjective-noun collocations. Firstly, it found some differences between how lemmatised collocation frequency counts predicted L1 English and Turkish speakers' RTs for adjective-noun collocations. The results showed that as the lemmatised collocation frequency counts increase, L1 Turkish participants need more time to respond to Turkish collocations. In contrast, as the lemmatised collocation frequency counts increase, L1 English participants need less time to respond to English collocations. That is to say, lemmatised collocation frequency counts predicted the RTs for English collocations in the same direction as non-lemmatised collocation frequency counts. However, lemmatised collocation frequency counts predicted the RTs for Turkish collocations in the opposite direction as non-lemmatised frequency counts. This finding is not totally unexpected from the corpus linguistics perspective. Although lemmatised collocations' LD-scores showed that agglutinating structure of Turkish appears to considerably increase adjective-noun pairs' collocational strength in Turkish, some of the high-frequency lemmatised collocations have lower LD and MI-scores than their base (unlemmatised) forms. A plausible explanation for the lower association scores of the some of the lemmatised collocations is that when the Turkish adjective-noun collocations are lemmatised, both the noun and collocation frequency increase with the addition of the inflected forms, but if the inflected forms do not collocate with the adjective to a similar degree as the base form, their collocational strength becomes weaker than the base (unlemmatised) forms (see section 4.3.1 for a more detailed discussion about the effect of agglutinating structure of Turkish on collocational strength). It should also be noted that some lemmatised high-frequency collocations used in this experiment have lower association scores than their unlemmatised forms (e.g. *ertesi gün*). Thus, lemmatised collocation frequency counts appeared to affect the L1 English and Turkish speakers' processing speeds of collocations differently.

The differences between how lemmatised and unlemmatised collocational frequency scores predicted L1 English and Turkish speakers' RTs to the collocations should be seen as tentative findings. Since the current study has not been specifically designed to compare the processing speeds of collocations whose lemmatised forms have lower scores of collocational strengths than the collocations whose lemmatised forms have higher scores of collocational strengths in Turkish, the results should be considered as tentative rather than conclusive. Furthermore, it would be fruitful to discuss the differing effects of lemmatised and non-lemmatised collocation frequency counts on the processing of collocations found in this study with other psycholinguistic experiments which look at the processing of MWS in other languages through the lens of lemmatised and non-lemmatised collocational (phrasal) frequency counts. However, (to the best of my knowledge) no experimental work seems to have explored the effect of lemmatised frequency counts on the processing of collocations in different languages yet. It should be noted that it is particularly important to look at the effects of lemmatised collocational frequency counts on the processing speeds of collocations in agglutinating languages since, this way it is possible to investigate the effect of agglutinating on the processing of collocations and other type of MWS. A methodological implication of this finding is that Turkish corpus and psycholinguistic works need to take account of the differences between lemmatised and unlemmatised forms of collocations, and other type of MWS. As shown in this study, researchers' decision about using lemmatised or unlemmatised collocations frequency and association counts in agglutinating languages is of paramount importance since they directly affect the findings of these studies and also the insights into the processing of MWS they provide.

It should be noted that only a few studies investigated the psycholinguistic reality of MWS in agglutinating languages. Durrant (2013) conducted a corpus-based exploration of within-words formulaic patterns in Turkish. He suggested high-frequency morphological patterns can be found within Turkish words, which provided evidence that formulaic patterning is not limited to word-level in agglutinating languages (see section 3.3 for a detailed review of Durrant's study). Furthermore, Cangir, Büyükkantarcioglu, and Durrant (2017) investigated whether there is any evidence of collocational priming in Turkish collocations. They found evidence for collocational priming for Turkish adjective-noun and noun-verb collocations. In addition to these findings, this study provided evidence that speakers of typologically different languages are equally sensitive to collocational frequency counts. The present study also identified some important differences between L1 Turkish and English speakers' sensitivity to single-word level frequency counts while processing adjective-noun collocations. However, the present study raises more questions than it resolves. First, this study cannot provide any further clarification on the extent to which L1 Turkish speakers process adjective-noun collocations without accessing to the individual words. To be able to address this question, a research design which specifically focusses on the activation of individual words is needed. Second, the current study does not explore the possibly differential effect of exclusivity of collocations in their processing in agglutinating versus non-agglutinating languages. Third, the present study does not investigate how inflected forms of the collocations are processed in agglutinating languages. In an agglutinating language, it is crucially important to find out if inflected forms of the collocations also enjoy a processing facilitation. Finally, the current study is not able to provide much insights into how the formulaicity between words interact with formulaicity within words in agglutinating languages (see section 7.4 for directions for future research addressing some of these research gaps).

7.2 Similarities and differences between L1 and L2 collocational processing

In addition to L1 processing of adjective-noun collocations in English and Turkish, this thesis explored L1 and advanced level L2 speakers' processing of adjective-noun collocations in English. The results showed that both L1 English and L2 English advanced groups processed the high-frequency collocations faster than the (non-collocational) baseline items, and low-frequency collocations faster than the baseline items in English. This finding suggests that L1 Turkish-English L2 advanced speakers are sensitive to the collocation frequency counts like L1 speakers of English because as the collocation frequency counts increased, L1 Turkish-English L2 speakers responded to the collocations in English more quickly, as L1 English speakers did. These findings add to the growing body of empirical research that advanced level L2 speakers are sensitive to the frequency distribution of frequently occurring MWS, just as L1 speakers are (Siyanova-Chanturia et al, 2011; Wolter & Gylstad, 2013; Wolter & Yamashita, 2018; Yi, 2018). The current study also looked at L1 English and L2 English advanced speakers' attending to single-word frequency counts alongside collocation frequency counts. The results indicated that both groups showed sensitivity to noun frequency counts, and L2 English advanced speakers did not appear to rely on the noun frequency scores more heavily than the L1 English group while processing adjective-noun collocations. That is to say, as the noun frequency increased, both groups needed less time to respond. Similar results have been reported by Wolter and Yamashita (2018). They found that both L1 and L2 speakers of English attend to word-level frequency and collocational frequency counts simultaneously.

It has been argued by some researchers that, most notably Wray (2002, 2008), that L1 (native) speakers and L2 (non-native) speakers process MWS in qualitatively different ways. That is, L1 speakers rely heavily on meaning assigned to larger chunks and L2 speakers rely heavily on words and rules. Wray (2002, 2008) suggests that L1 speakers process the MWS as single unit of meaning, whereas L2 speakers firstly decompose MWS into individual words and then process them word-by-word through semantic value of the individual words Wray (2002) provided the collocation *major catastrophe* as an example and claimed that L1 speakers process this as a single unit of meaning, but L2 speakers decompose it into its individual words as *major* and *catastrophe* (see section 1.3 for a discussion about Wray's approach). This view received some criticisms from empirical studies focussing on learning and processing of collocations. For example, Durrant and Schmitt (2010) found a learning effect for collocations presented in a training session, and they suggest that it is the lack of exposure to L2 input that explained the differences between L1 and L2 collocational processing. In line with this, Wolter and Yamashita (2018) found that both L1 and L2 speakers of English showed sensitivity to single word and collocational frequency counts. Furthermore, the current study replicated the findings of Wolter and Yamashita (2018) that both L1 and L2 advanced speakers of English attend to single-word and collocational frequency counts simultaneously. These findings largely conflict with Wray's (2002, 2008) views that L1 and L2 apeakers process MWS in fundamentally different ways. The present study and some other recent experiments (e.g. Siyanova-Chanturia et al. 2011; Wolter and Gyllstad, 2018) suggest a more unified approach in which all language users (including L2 speakers with a certain level of proficiency) are sensitive to the frequency information at different grain sizes (both single-word, and multi-word levels). These findings are in line with usage-based models of language (Ellis, 2002; Christiansen & Chater, 2016; Kemmer & Barlow, 2000; Tomasello, 2003), according to which frequency, and probability of input plays a key role in language acquisition and processing.

One of the under-researched topics in the field of processing of MWS is that the effects of collocational strength, also known or transitional probabilities and word-to-word contingency statistics on the processing of collocations by L1 and L2 speakers. Unfortunately, the current study has not addressed this important topic, and only a few studies have investigated L1 and L2 speakers' sensitivity to transitional probabilities during the processing of MWS. For example, McDonald and Shillcock (2003) found that L1 English speakers were sensitive to the transitional probabilities of the verb-noun collocations. However, Frisson, Rayner, and Pickering (2005) replicating and expanding this research, claimed that transitional probabilities have no significant effect on collocational processing if contextual predictability is controlled. Nevertheless, they argued that contextual predictability (as measured by cloze tests) involves some aspects of transitional probabilities so they do not entirely dismiss their effects on language processing. Nevertheless, Ellis et al. (2008) found that L1 speakers' processing is affected by the MI-scores of the MWS, whereas advanced L2 speakers' processing appear to be only affected by the phrasal frequency counts of MWS (see section 3.1 for a detailed discussion on MI-score and other AMs). Such findings are interesting, but questionable because Ellis et al. (2008) had a quite small sample size and they do not seem to have control over confounding variables such as single word and phrasal frequency counts of the MWS used in the experiments. Conversely, more recent studies, for example, Ellis et al. (2014) found that L2 speakers are sensitive to transitional probabilities in verb-argument constructions as measured by Delta P statistics. In line with this, Yi (2018) found that L2 are sensitive to MI-scores of the collocations. Based on these contrasting results from the previous studies, it remains unclear whether L2 speakers are sensitive to transitional probabilities within MWS. Another important question from a methodological perspective remains to be investigated is that the extent to which the type of association measure (e.g. MI, Delta P, LD) used to examine speakers' sensitivity to transitional probabilities affect the results of the experimental works with regard to L1 and L2 speakers' sensitivity to transitional probabilities (see section 7.4 for directions for future research about measuring transitional probabilities).

7.3 Methodological implications of this dissertation

Although psycholinguistics and corpus linguistics have quite different research goals, recently a number of studies that combined corpus and psycholinguistics methods tend to increase (e.g. Monaghan & Mattock, 2011; Millar, 2011; Reali, 2014). As mentioned previously (see section 2.1), for a fruitful combination of corpus and experimental methods, a special attention needs to be paid to the advantages and disadvantages of corpus and experimental methodology over the other, and in what ways they can complement each other. In this study, these two methods have complemented each other in two ways. First, the corpus study (see chapter 4) has made it possible to explore frequency and collocational strengths of adjective-noun collocations in the two languages. Looking at the corpus findings regarding frequency and association scales of collocations in English and Turkish, it was possible to generate hypotheses about the differences of mental processing of collocations in English and Turkish by L1 speakers of each language. Second, the general corpora of the two languages, the BNC and the TNC, were used as a source of experimental stimuli for the experiments designed to examine L1 English and Turkish speakers' processing of collocations in their first languages, and L1 an L2 processing of collocations in English (see section 6.1.2 for a detailed explanation on the preparation of experimental stimuli).

The first way this study combined the corpus and experimental methods involved drawing on both corpus and experimental data to gather evidence for the potential differences in L1 speakers' processing of collocations in English and Turkish. The main goal of the corpus study (see chapter 4) was to find out whether there was a considerable difference between the formulaicity of adjective-noun collocations in English and Turkish. It should be noted that there is no single method that can be used to compare the formulaicity of certain type of constructions in different languages. Therefore, this study looked at the frequency and collocational strength of the translation equivalent of adjective-noun collocations in English and Turkish. The reasons for focussing on the variables, frequency and collocational strength are that, first they are important features of collocations (see section 3.1 for a detailed discussion on collocational properties) second, they are likely to affect the processing of collocations (see Sonbul, 2015; Wolter & Yamashita, 2018; Yi 2018). In this regard, one advantage of conducting the corpus study before designing the psycholinguistic experiment was that the results of the corpus study could shed light on the hypothesis of the psycholinguistic experiment regarding potential differences in adjective-noun collocations' processing in English and Turkish. Another advantage of carrying out the corpus study before the experiment was that the corpus study made it possible to investigate the effect agglutinating structure of Turkish on the collocability of adjectives and nouns. In order to test the effect of agglutinating structure of Turkish, the corpus study looked at both unlemmatised and lemmatised adjective-noun combinations in English and Turkish. With lemmatising the extracted adjective-noun combinations, it was possible to see the overall effect of inflected and the base forms on the collocability of adjectives and nouns in the two languages. However, as Durrant (2013) also noted, lemmatisation needs to be treated with caution because its indiscriminate use inevitably views the base and inflected forms of the adjective-noun combinations as homogenous phenomena and conceals the separate collocational networks of the base and each inflected forms of the adjective-noun combinations.

This type of multi-method perspectives has also been employed by a few previous studies (e.g. Reali, 2014; Siyanova & Schmitt, 2008). For example, Siyanova and Schmitt (2008) investigated L2 learners' collocational productions and on-line processing by combining corpus and experimental approaches. They firstly explored the differences between L2 learners and L1 speakers' productions of adjective-noun collocations in terms of frequency and collocational strength. Their corpus study demonstrated that the vast majority of the collocations produced by L2 learners were native-like adjective-noun collocations. Based on this finding, Siyanova and Schmitt (2008) made predictions about the L1 and L2 processing of collocations. The experiments did not seem to fully confirm the findings of the corpus study since L2 learners did not seem to have native-like intuitions of collocations. A similar combined use of corpus and experimental approaches provided by Reali (2014), to investigate the distributional patterns and the L1 processing of Spanish relative clauses. More precisely, Reali (2014) firstly investigated whether L1 Spanish speakers have a preferred word order for subject and object relative clauses, using a spoken corpus of Spanish. She found that object relative clauses tend to be pronominal, while subject relative clauses tend to have a full noun phase in the embedded position. Based on these corpus findings, the predictions were made for the L1 processing of subject and object relative clauses in Spanish. The behavioural experiments largely mirrored the corpus findings in terms of processing of relative clauses in Spanish (see also section 2.1.1 for a detailed review of studies that combined corpus and experimental perspectives).

At this point, Durrant's (2013) corpus analysis on the complex inflectional patterns in Turkish is also worth mentioning since it is a seminal study providing a corpus-based description of formulaic patterns that are found within Turkish words. Durrant's (2013) study aimed to develop a fine-grained model of formulaic patterns that are found within agglutinating morphology of Turkish. Durrant (2013) suggests that formulaic patterning is not limited with word level in Turkish. More specifically, he found that most high-frequency morphemes enter into collocational relations with their syntagmatic neighbours. A number of high-frequency morpheme combinations are used across verb roots. Moreover, they also form strong relationships with particular verb roots. With regard to combining corpus and experimental methods, Durrant (2013) suggests that hypotheses about these types of formulaic patterns should not be based on psycholinguistic data alone, the corpus-based descriptions are needed for a fine-grained model of formulaic patterns found in complex word forms in agglutinating languages. In other words, without a thorough corpus-based description of those patterns, studies with psycholinguistic perspectives alone might adopt a dualistic paradigm in which complex word forms are either stored holistically or fully processed, (see Lehtonen & Laine, 2003; Gürel, 1999, for examples of studies which adopted the dualistic paradigm). The view of holistic storage versus full analysis might be too simplistic to capture the formulaicity within complex word forms in agglutinating languages (Durrant, 2013). Along these lines, the present thesis also provided an initial corpus analysis of adjective-noun collocations before setting up the psycholinguistic experiments to observe how nominal inflections affect collocability of adjectives and nouns in Turkish. Based on the findings of the corpus study, predictions were made about processing of English and Turkish adjective-noun collocations by L1 speakers of each language.

The second way this study combined corpus and psycholinguistic perspectives is that the general corpora of English and Turkish, the BNC and TNC were used as a source of experimental stimuli. Undoubtedly, using general and balanced corpora, the BNC and TNC was necessary to make sure that classifications of the English and Turkish items as high, low-frequency collocations and baseline were accurate. Therefore, all items' collocational frequency and LD-scores were gathered from the BNC and TNC. In addition, the items' single word frequency counts for adjectives and nouns were also extracted from the two corpora to closely match them across the conditions (see Table 6.3 and

Table 6.4, see also section 6.1.2 for a detailed explanations of item development). The baseline items were checked against the BNC and TNC to make sure they have no attestation in the corpora. Similar procedures were also followed by the previous psycholinguistic experiments looking at the processing of collocations and various types of MWS (e.g. Sonbul 2015; Vilkaite, 2016; Wolter & Yamashita, 2018). For example, Wolter and Yamashita (2018) investigated the effects of congruency, collocational frequency, and single-word frequency counts on collocational processing for L1 and L2 speakers. Therefore, their study included four types of items, congruent collocations, incongruent collocations, translated Japanese-only collocations to test if they are activated for Japanese L1-English L2 speakers when they process collocations in English, and the baseline items. They checked all the items against the Contemporary Corpus of American English (COCA) to make sure that their classifications were accurate. That is to say, congruent and incongruent collocations frequently occur, and Japanese only and baseline items were not found in the COCA.

There is an important difference between the way that the current study and previous experiments operationalised collocations. The vast majority of the previous experimental studies used MI-scores to operationalise collocations (e.g. Siyanova & Schmitt, 2008; Ellis et al. 2009; Vilkaite, 2016; Wolter & Yamashita, 2015). As also discussed previously (see section 3.1), the MI-score is not without its drawbacks (see also Gablasova et al. 2017). Importantly it should be noted that the MI-score does not only highlight the exclusive collocations but it has also been found to favour low-frequency collocations (e.g. Granger & Bestgen, 2014). Therefore, it may not be the most suitable measure for extracting high-frequency collocations. Furthermore, one important drawback of the MI score is that it is based on a logarithmic scale to express the ratio between the frequency of the collocation, and the frequency of the random co-occurrence of the word combinations (Church & Hanks, 1990). This random co-occurrence approach is similar to considering corpus as a box in which all words are written in small papers, and then it is thoroughly shaken. It is questionable that whether this random co-occurrence approach is a reliable baseline to extract collocations (Gablasova et al. 2017). Another important drawback of MI-score is that although it is a normalised scale, it does not have theoretical maximum and minimum values. Therefore, it may not always be easy to interpret the MI-scores and compare them across corpora. Considering these drawbacks of the MI-score, the current study used LD-score to operationalise the adjective-noun collocations. The LD-score is based on the harmonic mean of two proportions that express the tendency of two words to co-occur relative to the frequency of these words in the corpus. Therefore, it avoids the potentially problematic shake-the-box approach (Gablasova et al. 2017), that is the random distribution model of language because it does not include the expected frequency on its equation (see appendix A for the equations of AMs). Practically, the LD-score highlights exclusive word combinations but not necessarily the low-frequency combinations. Another advantage of using the LD-score is that it operates on a clear delimited scale with a maximum value of 14. Therefore, the LD-score is directly comparable across corpora.

Alongside the many benefits of combining corpus and experimental methods, it is also important to discuss the challenges that was experienced in combining the two methods. Most of the challenges arouse from the differences between CL and psycholinguistics' methodological choices related to the use of frequency and AMs. For example, it is perfectly acceptable to use normalised frequency scores to per million words in corpus linguistics as a measure of frequency, however, it is more preferable to use log transformed versions of the frequency scores in psycholinguistics (e.g. Wolter & Yamashita, 2018). Therefore, for the experimental component of the thesis, the single word and collocational frequency counts were log transformed using the SUBTLEX Zipf scale (Van Heuven et al. 2014). However, relative frequency scores were used in the corpus component. Similarly, CL and psycholinguistics studies also have slightly different approaches to collocational AMs. Psycholinguistics studies seem to rely on the MI-score predominantly to operationalise collocations (Wolter & Yamashita, 2017; Yi, 2018), while studies with corpus linguistics focus seems to be more motivated to explore various AMs such as LD, and Delta P (e.g. Gries, 2013; Gablasova et al. 2017; Kang, 2018). Unlike previous psycholinguistics experiments, the current study used the LD-score, which seems to be more standardised and systematic than the MI-score considering its mathematical reasoning and the scale on which it operates.

A challenge associated with choosing the LD-score was determining the LD-score based threshold values for operationalising collocations. Conventionally psycholinguistic studies employing the MI-score to extract collocations used the MI-score of 3 as a threshold for highlighting the exclusive co- collocations (Wolter & Yamashita, 2015; Yi, 2018). Since no experimental work has employed the LD measure so far, no conventional threshold score has been determined yet. In this case, the current study had to determine threshold LD-scores for extracting high- and low-frequency collocations. Considering the frequency distribution of the LD-score based collocations in the corpus study, this study chose LD-score of 7 as a threshold for high-frequency collocations, and LD scores of 2 and 4 as a threshold for low-frequency collocations (both scores within a 3-3 collocation window span). The main question was to determine these threshold values were what scores of LD correspond to high- and low-frequency collocations. Looking at the results of the experiments, this approach seems to have worked. Nevertheless, it should be noted that this way of determining the threshold values is far from being perfect since it does not involve a standardisation study in which the RTs are compared for collocations extracted from various frequency and LD-score bands. This is one of the research gaps in the field to be addressed by the future research (see also section 7.5.1

for a further discussion on this topic). The next section summarises the limitations of this study and provides some directions for future research.

7.4 Limitations and directions for future research

As also mentioned previously (see sections 4.3.3 and 6.4.3), this study is not without limitations. One of the most important limitation for the corpus study is that it only included the congruent adjective-noun combinations that have translation equivalents in both English and Turkish. Looking only at congruent word combinations enabled the corpus study (see chapter 4) to compare the frequency counts and AMs of the items that they have very similar meanings in English and Turkish. However, it would also be useful to observe how the randomly selected congruent and incongruent adjective-noun combinations affect the scales of frequency and associations in the two languages. Therefore, a future corpus study examining the formulaicity of English and Turkish collocations contrastively might include the incongruent word combinations that do not have translation equivalents in the both languages alongside the congruent ones. Methodologically, an ideal way of including both congruent and incongruent collocations would be to randomly select a certain number of word combinations from the collocational lists of each node words in English and Turkish. Then, it would be possible to create a scale of frequency counts and AMs of lemmatised and unlemmatised collocations that are randomly selected from each frequency bands in the two languages. This way the selected items of corpus analysis would include both congruent and incongruent collocations and it would be possible to compare the frequency and association scales of randomly selected items in each frequency bands in English and Turkish.

One of the most important limitation of the experimental study (see chapter 6) is that it did not include a priming LDT to investigate the potential differences of priming effects between adjective-noun collocations in Turkish and English. Therefore, future psycholinguistic experiments should employ a lexical or semantic priming paradigm to examine the extent to the priming effects between Turkish and English adjectives and nouns of high-frequency collocations differ from each other. This type of design is particularly important for addressing the question that whether Turkish collocations tend to be processed more holistically than English collocations. As Siyanova-Chanturia (2015) also noted, to examine the extent to which the individual words play a role in the processing of the whole MWS, or explore the effect of single word frequency relative to the frequency of the whole MWS, a semantic or syntactic priming paradigm should be employed (p. 287). Another important limitation of the current study is that it did not address whether the inflected forms of the adjective-noun collocations also enjoy processing facilitation, like their base forms do. Future experiments should address this question by including the inflected collocations alongside the base forms of and baseline items. One difficulty of designing this type of experiment is related to matching the frequencies for the single words and collocations across base and inflected conditions since the inflected forms of the nouns generally tend to have lower frequency counts than their base forms.

In addition to the limitations of the current study, it is also important to discuss the future research possibilities to investigate the formulaicity of agglutinating languages from CL and psycholinguistic perspectives. One possibility is to examine the psycholinguistic reality of high-frequency morphological patterns found within Turkish words. Carrying out a corpus analysis, Durrant (2013) found that high-frequency morphological patterns exist within Turkish words, which provided evidence that formulaic patterning is not limited to word-level in agglutinating languages. To the best of my knowledge, no experimental study has investigated yet whether these identified high-frequency morpheme combinations are psycholinguistically real for L1 and L2 speakers of Turkish. The findings of such an experimental work would provide empirical evidence that whether Turkish speakers are sensitive to the high-frequency morpheme combinations within Turkish words. It should be noted that one of the crucial methodological decision that this type of study needs to make related to choosing the suitable control items. It would probably be appropriate to compare the processing speeds of the words with high-frequency morpheme combinations with the words with low-frequency morpheme combinations and words without any inflections. Undoubtedly, closely matching the word and constituent frequencies of the items in all of the three conditions is crucial for the robustness of the experiment.

Processing of lexical bundles in English and Turkish is another interesting research possibility to be investigated using both CL and psycholinguistics methods. As Durrant (2013) noted, meanings that require multiple word combinations in English, can be expressed using single inflected words in Turkish. That is to say, individual word forms on average have considerably lower frequency counts than their English equivalents. Since the individual words have lower frequency counts, so too are three- or four words lexical bundles than their equivalents in English. Definitely, this prediction needs to be tested through a thorough corpus investigation of English and Turkish lexical bundles contrastively. Furthermore, this corpus study could also investigate the proportion of the three- four-word lexical bundles in English and Turkish. If the corpus analysis confirm the predictions that three- four-word lexical bundles occur less frequently in Turkish than in English and the proportion of the bundles in Turkish is lower than in English, it would also be interesting to examine the processing differences of English and Turkish lexical bundles by L1 and L2 speakers of each language. The psycholinguistic experiment should mainly focus on the possible differences between L1 Turkish and English speakers' attending to single word and phrasal frequency counts of the bundles during the processing of the items. This type of empirical studies that combine corpus and experimental methods potentially play an important role in identifying the differences between formulaicity of agglutinating and non-agglutinating languages.

As discussed previously (see section 7.3), the effects of collocational strength, also known as transitional probabilities, in L1 and L2 collocational processing is an under-researched topic. So far only a few studies have investigated L1 and L2 speakers' sensitivity to collocational strength during the processing of collocations. The studies looked at the effect of collocational strength so far have reached conflicting results. For example, McDonald and Shillcock (2003) found that L1 English speakers were sensitive to the transitional probabilities of the verb-noun collocations, as measured by forward and backward measures of transitional probabilities. However, Frisson, et al. (2005) claimed that transitional probabilities have no significant effect on collocational processing. More recent studies, for example, Ellis et al. (2014) found that L2 speakers are sensitive to transitional probabilities in verb-argument constructions as measured by Delta P statistics. In line with this, Yi (2018) found that L2 are sensitive to MI-scores of the collocations. Based on these contrasting results from the previous studies, it remains unclear whether L2 speakers are sensitive to transitional probabilities within MWS. As can be seen, the experimental studies investigated the effects of collocational strength use different AMs to operationalise collocational strength such as MI, Delta P. As Gablasova et al (2017) also noted, the AMs used in these studies are likely to affect the results since they highlight the different aspects of the collocational associations (see section 3.1 for a review of AMs). It is therefore important to address the research gap that the AMs used for operationalising the collocations and measuring the effect of transitional probabilities need to be validated and standardised through combining experimental data and corpus data. In order to do that the first step would involve a corpus study to identify what type collocations each measure would highlight in terms of single word and collocational frequency. The second step involves the comparisons of the processing times of collocations that different AMs highlighted.

Another interesting research possibility for studies is investigating the individual differences in the processing of MWS. It is important to bear in mind that individual differences can particularly have large impact on L2 learning processes (see DeKeyser & Koeth, 2011; Pawlak, 2017 for reviews). Therefore, L2 processing of collocations are also likely to be affected by the speakers' individual differences. Some individual differences have particularly good explanatory power in the processing of MWS. For example, phonological short-term memory capacity (Bolibaugh & Foster, 2013), and working memory capacity (Tremblay & Baayen, 2010) affect the L2 processing of MWS. To the best of my knowledge, the effect of declerative and procedural memory capacities on the L1 and L2 processing of MWS has not been explored yet. The effect of declarative memory on the processing of lexical and grammatical aspects is expected to be different (Ullman, 2015). With respect to vocabulary, the acquisition of lexical information in both first and second language occurs in the declarative memory system. Therefore, declarative memory capacity might affect the processing of MWS. Procedural memory is also likely to play an important role in the processing of MWS since advanced command of MWS is less likely to require conscious processes (Bolibaugh & Foster, 2013). These predictions need to be tested through experimental studies to understand the underlying mechanisms of L1 and L2 processing MWS better.

There are also interesting research possibilities from a learner corpus research perspective. For example, to the best of my knowledge only one study has looked at the L2 learners' use of collocations or other type of MWS over time from a longitudinal perspective (see Bestgen &

Granger, 2014). To be able to observe the development of a group of L2 learners' use of MWS, a learner corpus that is designed to include longitudinal data is needed. An important advantage of longitudinal corpus studies is that it is possible to trace the development of the same individual learners over a period of time in terms of the frequency and associations of the collocations they use. This way it is also possible to observe the individual learner differences with regard to the use of collocations. Another learner corpus research possibility is to observe the effect of the tasks such as monologic versus dialogic tasks on the L2 learners' use of collocations and other type of MWS. To be able to observe the effects of the type of tasks on the use of collocations, a learner corpus that is designed to include both monologic and dialogic tasks is needed. This type of learner corpus studies has a potential to produce pedagogical implications with respect to teaching collocations in EFL and ESL classrooms. It is noteworthy that these suggestions about future research on MWS does not intend to be a comprehensive or complete list, it intends to encourage further research on the processing and use of MWS in typologically different languages and combining corpus and experimental methodologies.

7.5 Concluding remarks

To conclude this thesis, I would like to argue that the combination of corpus data with psycholinguistic experimentation enables the researchers to approach the phenomena of MWS from two different angles, and it also represents a useful synergy. The combination of these methodologies is still not a very common approach (see also section 2.1 for a discussion on the combination of these methods), so that some review studies presented surveys of different ways in which corpus data and experimentation can be combined (e.g. Durrant & Siyanova-Chanturia, 2015; Rebuschat et al. 2017). I hope this thesis demonstrates that integration of corpus data and psycholinguistic experimentation is both doable and at the same time an empirically robust approach to study the MWS in typologically different languages. When it comes to the main findings of this thesis, L1 English and Turkish speakers appear to process adjective-noun collocations that have similar frequency and collocational strength counts in English and Turkish equally quickly in their respective languages. Although they process collocations equally quickly, L1 English and Turkish speakers differ in their attendance to single word frequency counts of nouns. The variation in L1 English and Turkish speakers' attending to single word frequency counts can be attributed to the typological difference of the two languages. This thesis also found that L2 English advanced speakers are sensitive to the collocation frequency counts like L1 speakers of English because as the collocation frequency counts increased, L2 English speakers responded to the collocations more quickly, as L1 English speakers did. Furthermore, both groups showed sensitivity to noun frequency counts alongside the collocational frequency counts. These finding add support to the growing body of empirical evidence that usage-based approaches to language acquisition (Bybee, 1998; Ellis, 2002; Christiansen & Chater, 2016a; Kemmer & Barlow, 2000; McCauley & Christiansen, 2016), which view linguistic productivity as a gradually emerging process of storing and abstracting MWS.

References

Aksan, Y., Aksan, M., Koltuksuz., A, Sezer, T., Mersinli, Ü., Demirhan, U., … Kurtoglu Ö. (2012). Construction of the Turkish National Corpus (TNC). *Proceedings of the 8th International Conference on Language Resources and Evaluation* (pp. 3223-3226). Istanbul, Turkey.

Altmann, G. T. M., & Kamide, Y. (1999) Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition 73,* 247–264.

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word combinations. In A.P. Cowie (ed.), *Phraseology. Theory, Analysis, and Applications* (pp. 101-122). Oxford: Oxford University Press.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62,* 67–82.

Arnon, I., & Cohen Priva, U. (2015). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *Mental Lexicon 9,* 377–400.

Aston, G., & Burnard, L. (1998). *The BNC Handbook: Exploring the British National Corpus with SARA.* Edinburgh: Edinburgh University Press.

Badecker, W. (2001). Lexical composition and the production of compounds: Evidence from errors in naming. *Language and Cognitive Processes 16,* 337–366

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning. *Psychological Science, 19,* 241–248.

Barr. D. J., Levy, R., Scheepers, C., & Tily, H., J. (2013). Random effects structure for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language, 68,* 752-793.

Barton, K. (2016). MuMIn: Multi-model inference. R package version 1.15.6. Retrieved from https://CRAN.R- project.org/package5MuMIn.

Bartsch, S., & Evert, S. (2014). Towards a Firthian notion of collocation. *Network Strategies, Access Structures and Automatic Extraction of Lexicographical Information.* 2nd Work Report of the Academic Network Internet Lexicography, OPAL–Online publizierte Arbeiten zur Linguistik. Mannheim: Institut fur Deutsche Sprache.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412.

Benor, S., & Levy, R. (2006). The chicken or the egg? A probabilistic analysis of English binomials. *Linguistic Society of America*, 82, 233–278.

Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing 26,* 28–41.

Brezina, V., McEnery, T., & Wattam, S. (2015). Collocations in context: a new perspective on collocation networks. *International Journal of Corpus Linguistics 20*, 139-173.

Brezina, V. (2019). *Statistics in Corpus Linguistics: A practical guide.* Cambridge: Cambridge University Press.

Biber, D., & Conrad, S. (2003). The frequency and use of lexical bundles in conversation and academic prose. *Lexicographia: International Annual for Lexicography,* 20, 56-71

Biber, D., Conrad, S., & Cortes, V. (2004). 'If you look at . . . lexical bundles in academic lectures and textbooks'. *Applied Linguistics, 25,* 371–405.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing. *International Journal of Corpus Linguistics, 14,* 275–311.

Bod, R. (2001). Sentence memory: Storage vs. computation of frequent sentences. Paper presented at the CUNY 2001, University of Pennsylvania, Philadelphia, PA

Bod, R. (2006). Exemplar-based syntax: How to get productivity from exemplars. *The Linguistic Review 23,* 291–320.

Bolibaugh, C., & Foster, P. (2013). Memory-based aptitude for nativelike selection: the role of phonological short term memory. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment.* (Vol. 35, pp. 205). (Language Learning & Language Teaching). Amsterdam: John Benjamins Publishing Company.

Breheny, P., & Burchett, W. (2017). Visualization of Regression Models Using visreg. *The R Journal, 9,* 56-71.

Buerki, A. (2016). Formulaic sequences: A drop in the ocean of constructions or something more significant? European Journal of English Studies, 20(1), 15-34.

Bybee, J. (1998). The emergent lexicon. *Chicago Linguistic Society, 34,* 421–435.

Bybee, J. & McClelland, J. L. (2005) Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. In N. Ritter (Ed.), T*he role of linguistics in Cognitive Science.* Special Issue of The Linguistic Review, 22 (2-4), (pp. 381–410).

Cangir, H., Büyükkantarcıoğlu. N., & Durrant P. (2017). Investigating collocational priming in Turkish. *Journal of Language and Linguistic Studies, 13*, 465-486.

Carrol, G., & Conklin, K. (2017). Cross language priming extends to formulaic units: Evidence from eyetracking suggests that this idea "has legs.*" Bilingualism: Language and Cognition,* 20, 299–317.

Carrol, G., & Conklin, K. (2019). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech,* 1-28

Cheng, W., Greaves, C., Sinclair, J., & Warren, M. (2009) Uncovering the extent of the phraseological tendency: Towards a phraseological analysis of concgrams. *Applied Linguistics*. *30,* 236-252

Chomsky, N. (1965). *Aspects of the theory of syntax*. Oxford, England: M.I.T. Press.

Christiansen, M. H., & Arnon, I. (2017). More than words: the role of multiword sequences in language learning and use. *Topics in Cognitive Science, 9,* 542-551.

Christiansen, M. H., & Chater, N. (2016a). *Creating language: Integrating, evolution, acquisition and processing*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2016b). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioural & Brain Sciences, 39,* e62.

Church, K. W., & Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Computational Linguistics*, 16(1), 22–29.

Colson, J. (2008). Cross-linguistic phraseological studies: an overview. In S. Granger & F. Meunier (Eds.), *Phraseology: an interdisciplinary perspective* (pp. 191-206). Amsterdam: John Benjamins Publishing.

Conklin, K., & Carrol, G. (2016). Native and non-native sensitivity to word order in novel multiword sequences. Paper presented at BAAL Vocabulary SIG, Nottingham, UK.

Cortes, V. (2008). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora, 23,* 397-423.

Dayrell, C. (2007). A quantitative approach to compare collocational patterns in translated and non-translated texts. *International Journal of Corpus Linguistics, 12,* 375–414.

DeCock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learning English on computer* (pp. 67 79). London: Addison, Wesley, Longman.

DeKeyser, R. M., and Koeth, J. (2011). Cognitive aptitudes for second language learning. In E. Hinkel (Ed.), *Handbook of Research in Second Language Teaching and Learning*, Vol. II, (London: Routledge), 395–406.

Durrant, P., & Schmitt, N. (2009). To what extent do native and non-native writers make use of collocations? *IRAL-International Review of Applied Linguistics in Language Teaching, 47*, 157–177.

Durrant, P., & Schmitt, N. (2010). Adult learners' retention of collocations from exposure. *Second Language Research, 26,* 163–188.

Durrant, P., & Doherty, A. (2010). Are high frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory, 6,* 125-155

Durrant, P. (2013). Formulaicity in an agglutinating language: The case of Turkish. *Corpus Linguistics and Linguistic Theory, 9*, 1–38.

Durrant, P. (2014). Corpus frequency and second language learners' knowledge of collocations: A meta-analysis. *International Journal of Corpus Linguistics, 19,* 443–477

Durrant, P. & Siyanova-Chanturia. A. (2015). Learner corpora and psycholinguistic research. In S. Granger, G. Gilquin, & F. Meunier (eds.), *Cambridge handbook of learner corpus research.* Cambridge: Cambridge University Press.

Ebeling, S. O., & Hasselgård, H. (2015). Learner corpora and phraseology. In S. Granger, G. Gilquin, & F. Meunier (Eds.), *The Cambridge handbook of learner corpus research* (pp. 185–206). Cambridge, UK: Cambridge University Press.

Ellis, N. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition, 24,* 143–188.

Ellis, N. (2008). Usage-based and form-focused language acquisition: The associative learning of constructions, learned attention, and the limited L2 endstate. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 372–405). New York, NY, and London, UK: Routledge.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42,* 375–396.

Ellis, N., & Simpson-Vlach R. (2009). Formulaic language in native speakers: Triangulating psycholinguistics, corpus linguistics, and education. *Corpus Linguistics and Linguistic Theory, 5,* 61-78

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse, 20,* 29–62.

Evert, S. (2005). *The statistics of word co-occurrences: Word pairs and collocations.* (Doctoral dissertation, Institut fur maschinelle Sprachverarbeitung, University of Stuttgart.)

Firth, J. (1957). *Papers in Linguistics.* Oxford, UK: Oxford University Press.

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of Contextual Predictability and Transitional Probability on Eye Movements During Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 862-877.

Frost, R., Armstrong, B.C., Siegelman, N. & Christiansen, M.H. (2015). Domain generality vs. modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19,* 117-125.

Gablasova, D., Brezina, V., McEnery, T. (2017). Collocations in corpus-based language learning research: identifying, comparing and interpreting the evidence. *Language learning 67,* 155-177.

Gablasova, D., Brezina, V., McEnery, T. (2017). Exploring learner language through corpora: Comparing and interpreting corpus frequency information. *Language learning 67,* 130-154.

Gareth, C., & Conklin, K., (2019). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech,* 0, 1-28

Gesmundo, A., & Smardžić, T. (2012). Lemmatisation as a tagging task. *In Proceedings of the 50th annual meeting of the association for computational linguistics, 50*, 368-372.

Granger, S., & Meunier, F. (Eds.). (2008). *Phraseology: An interdisciplinary perspective.* Amsterdam: John Benjamins Publishing.

Granger, S. (2014). A lexical bundle approach to comparing languages: Stems in English and French. *Languages in contrast: international journal of contrastive linguistics, 14, 58-72*.

Gries, S. Th. (2008). Dispersions and adjusted frequencies in corpora. *International Journal of Corpus Linguistics, 13,* 403–437.

Gries, S. Th. (2013). 50-something years of work on collocations: What is or should be next . . . . *International Journal of Corpus Linguistics, 18,* 137–166.

Gilquin, G., & Gries, S. T. (2009). Corpora and experimental methods: A state-of-the-art review. *Corpus Linguistics and Linguistic Theory, 5,* 1–26.

González Fernández, B., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have?: The effects of frequency and amount of exposure. *International Journal of Applied Linguistics, 166,* 94-126.

Gómez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4,* 178–186.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language.* Oxford, UK: Oxford University Press.

Göksel, A., & Kerslake, C. (2005). *Turkish: A comprehensive grammar.* London: Routledge.

Gürel, A. (1999). Decomposition: To what extent? The case of Turkish. *Brain and Language, 68*, 218-224.

Hale, J. (2006). Uncertainity about the rest of the sentence. *Cognitive Science*, 30, 609-642.

Hardie, A. (2012). CQPweb—Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics, 17,* 380–409.

Hinojosa, J. A., Moreno, E. M., Casado, P., Munõz, F., & Pozo, M. A. (2005) Syntactic expectancy: An event-related potentials study. *Neuroscience Letters, 378,* 34–39.

Hoey, M. (1991). *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hoey, M. (2005). *Lexical priming: A new theory of words and language.* London: Routledge.

Hoff, E. (Ed.). (2011). Research methods in child language: A practical guide. Malden, MA: Wiley-Blackwell.

Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics, 19,* 24–44.

Hunston, S. (2002). *Corpora in applied linguistics.* Cambridge, UK: Cambridge University Press.

Jackendoff, R. (1997). *The architecture of the language faculty.* Cambridge, MA: MIT Press.

Janssen, N. & Barber H. A. (2012). Phrase frequency effects in language production. *Plos One*, 7, 1-11

Jiang, N. (2012). *Conducting reaction time research in second language studies.* London: Routledge

Jolsvai, H. McCauley, S.M. & Christiansen, M.H. (2013). Meaning overrides frequency in idiomatic and compositional multiword chunks. In M. Knauff, M. Pauen, N. Sebanz, &

I. Wachdmuth (Eds.), Proceedings of the 35th Annual Conference of the Cognitive Science Society, (pp. 692-697). Austin, TX: Cognitive Science Society.

Juhasz, B. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (eds.), *Eye movements: A window on mind and brain,* 374–389. Oxford: Elsevier Science.

Kang, B. (2018). Collocation and word association: comparing collocation association measuring methods. *International Journal of Corpus Linguistics, 24,* 85–113

Kemmer, S., & Barlow, M. (2000). Introduction: a usage-based conception of language. In M. Barlow & S. Kemmer (eds.), *Usage based models of language,* vii–xxviii. Stanford: CSLI Publications.

Kuperman, V., Schreuder R., Bertram, R., & Baayen. R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology 35,* 876–895.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests in linear mixed effects models. R package version 2.0-29. Retrieved from https://CRAN.Rproject.org/package5lmerTest.

Lenth, R. (2016). Least-square means: T R package lsmeans. *The Journal of Statistical Software, 69,* 1-33.

Lehtonen, M., & Laine, M. (2003). How word frequency affects morphological processing in monolinguals and bilinguals. *Bilingualism: Language and cognition 6,*. 213–225.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical tests for advanced learners f English. *Behaviour Research Methods, 44,* 325-343.

Libben, M., & Titone, D. (2008). The multidetermined nature of idiom processing. *Memory and Cognition,* 36, 1103–1121.

Lohmann, A. (2012). A processing view on order in reversible and irreversible binomials. *Vienna English Working Papers,* 21, 25–50.

Lowder M., & Gordon. P. C. (2016). Eye-tracking and corpus-based analyses of syntax-semantics interactions in complement coercion. *Language, Cognition and Neuroscience, 31,* 921-939.

Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11,* 122–134.

McDonald, S. A., & Shillcock, R. C. (2003). Eye-movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14,* 648–652.

MacWhinney, B. (2017). A shared platform for studying second language acquisition. *Language Learning*, *67*, 255–276.

McCauley, S. M., & Christiansen, M.H. (2017). Computational investigations of multiword chunks in language learning. *Topics in Cognitive Science, 9,* 637-652.

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice.* Cambridge, UK: Cambridge University Press.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review, 63,* 81-97.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics 32,* 129–148.

Misyak, J. B., & Christiansen, M.H. (2010). When more in statistical learning means less in language: Individual differences in predictive processing of adjacent dependencies. In R. Catrambone & S. Ohlsson (Eds.), Proceedings of the 32nd Annual Cognitive Science Conference (pp. 2686 - 2691). Austin, TX: Cognitive Science Society.

Misyak, J.B., Christiansen, M.H., & Tomblin, J.B. (2010). On-line individual differences in statistical predict language processing. *Frontiers in Psychology 1,* 1-9

Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes, 28,* 1–27.

Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition, 123,* 133-143

Monaghan, P., & Rowland, C. (2017). Combining language corpora with experimental and computational approaches for language acquisition research. *Language learning, 67,* 15-40

Mollin, S. (2012). Revisiting binomial order in English: Ordering constraints and reversibility. *English Language and Linguistics,* 16, 81–103.

Mondini, S., Luzzatti, C., Saletta, P., Allamano, N., & Semenza, C. (2005). Mental representation of prepositional compounds: Evidence from Italian agrammatical patients*. Brain and Language 94,* 178–187.

Morgan, E., & Levy, R. (2016). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition,* 157, 384–402

Myles, F., & Cordier, C. (2017). Formulaic sequence (FS) cannot be an umbrella term in SLA: Focussing on psycholinguistic FSs and their identification. *Studies in Second Language Acquisition. 39,* 3-28.

Naimi, B. (2015). USDM: Uncertainty analysis for species distribution models. R package version 1.1-15. Retrieved from https://CRAN.R-project.org/package5usdm.

Nesselhauf, N. (2005). *Collocations in a learner corpus.* Amsterdam: John Benjamins.

Nguyen, T. M. H., & Webb, S. (2016). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research,* 1–23. Advance online publication.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal 91,* 433–445.

O'Grady, W. (2005). *How children learn language.* Cambridge, UK: Cambridge University Press.

Pawlak, M. (2017). Overview of learner individual differences and their mediating effects on the process and outcome of L2 interaction. In L. Gurzynski-Weiss (Ed.), *Expanding Individual Difference Research in the Interaction Approach*, (pp. 19–40). Amsterdam: John Benjamins Publishing.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. Richards & R. Schmidt (Eds.), *Language and communication* (pp. 191–226). London, UK: Longman.

Perruchet, P. & Peereman, R. (2004). The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics, 17,* 97-119.

Pecina, P. (2010). Lexical association measures and collocation extraction. *Language Resources and Evaluation, 44 (1-2),* 137–158.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. (2019). PsychoPy2: Experiments in behaviour made easy. *Behaviour Research Methods, 51,* 195-203.

Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Basic Books.

Pinker, S. & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Science*, 6(11), 456-463

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Retrieved from: http://www.R-project.org

Rayson, P., & Garside, R. (1998). The CLAWS web tagger. *ICAME Journal, 22,* 121-123

Reali, F. (2014). Frequency affects object relative clause processing: Some evidence in favour of usage-based account. *Language learning, 64,* 685-714

Rebuschat, P. & Williams, J. N. (Eds.). (2012). Statistical learning and language acquisition. Berlin: Mouton de Gruyter

Rebuschat, P. Meurers, D., McEnery, T. (Eds.). (2017). Language learning research at the intersection of experimental, computational and corpus-based approaches: An introduction. *Special issue of Language learning, 67,* 6-13

Rowland, C. (2013). *Understanding child language acquisition.* London: Routledge.

Rychly, P. (2008). A lexicographer-friendly association score. In P. Sojka & A. Hor´ak (Eds.), Proceedings of recent advances in Slavonic natural language processing, RASLAN (pp. 6–9). Brno: Masaryk University

Russell L., V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69,* 1–33.

Saffran, J. R., Aslin, R. N., & Newport, E. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928

Schmitt, N. (2012). Formulaic language and collocation. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–10). New York: Blackwell.

Smadja, F., McKeown, K. R., & Hatzivassiloglou, V. (1996). Translating collocations for bilingual lexicons: A statistical approach. *Computational linguistics, 22,* 1–38.

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford, UK: Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, Corpus and Discourse.* London: Routledge.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian Modern Language Review, 64,* 429–458.

Siyanova-Chanturia, A. (2015). On the "holistic" nature of formulaic language. *Corpus Linguistics and Linguistic Theory, 11,* 285–301.

Siyanova-Chanturia, A. & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, *36,* 549-569

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 776–784.

Siyanova-Chanturia, A., Conklin, K., Caffara, S., Kaan, E., Walter J. B., van Heuven, W. J. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language, 175,* 111-122.

Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition, 18,* 419-437.

Staub, A. & Clifton, C. (2006) Syntactic prediction in language comprehension: Evidence from either…or. *Journal of Experimental Psychology: Learning Memory and Cognition*, 32(2), 425-436

Stubbs, M. (2009) The search for units of meaning: Sinclair on empirical semantics. *Applied Linguistics*, *30,* 115-137.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning 61,* 569-613.

Thompson, S. P., & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3,*1–42.

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual, monolingual idiom processing. In R. R. Heredia, & A. A. Cieslicka (Eds.), Bilingual figurative language processing (pp. 171-207). New York, NY: Cambridge University Press.

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (ed.), *Formulaic sequences,* 153–172. Amsterdam: John Benjamins Publishing.

Ullman, M. (2001). The declarative/procedural model of lexicon and grammar. *Journal of Psycholinguistic Research,* 30(1), 37-69.

Ullman, M. (2015). The declarative/procedural: A neurobiologically motivated theory of first and second language. In B. VanPatten, J. Williams (Eds.), Theories in Second Language Acquision (pp. 135-158). New York, NY: Taylor and Francis

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40,* 831-856.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005) Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition 31,* 443–467.

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). Subtlex-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology, 67,* 1176-1190

Vetchinnikova, S. (2014). *Second language and the idiom principle* (Doctoral dissertation, University of Helsinki)

Vilkaite, L. (2016). Are non-adjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 1632-1642.

Weinert, R. (2010). Formulaicity and usage-based language: linguistic, psycholinguistic and acquisitional manifestations. In D. Wood (ed.), Perspectives on formulaic language. Acquisition and communication. London/New York: Continuum. 1-20.

Williams, J. N. (2009). Implicit learning in second language acquisition. In William C. Ritchie and Tej K. Bhatia (Eds.), *The New Handbook of Second Language Acquisition*, Emerald Group Publishing Limited, pp. 319-353.

Wiechmann, D. (2008). "On the computation of collostruction strength: Testing measures of association as expressions of lexical bias". *Corpus Linguistics and Linguistic Theory, 4,* 253–290.

Williams, G. (1998). Collocational networks: Interlocking patterns of lexis in a corpus of plant biology research articles. *International Journal of Corpus Linguistics, 3,* 151–171.

Williams, J. N. (2012). Working memory. In S. Gass & A.Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 427-441). New York: Routledge

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge, Cambridge University Press.

Wray, A. (2008). Formulaic language: pushing the boundaries. Oxford: *Oxford University Press*

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics, 32,* 231-254

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics, 32,* 430–449.

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing. *Studies in Second Language Acquisition, 35,* 451–482.

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics, 36,* 1193–1221.

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition, 40,* 395–416

Wood, D. (2015). *Fundamentals of formulaic language: An introduction.* London/New York: Bloomsbury.

Wulff, S. (2009). Converging evidence from corpus and experimental data to capture idiomaticity. *Corpus Linguistics and Linguistic Theory, 5*, 131-159

Xiao, R. & McEnery, T. (2006). Collocation, semantic prosody, and near synomyny: A corpus-linguistic perspective. *Applied linguistics, 27(1),* 103–129

Appendices

Appendix A: Collocational Association measures

The t-score is calculated as follows:

$$\mathrm{T-\ score} = \frac{O_{11} - \frac{R_1 \times C_1}{N}}{\sqrt{O_{11}}}$$

The MI-score is calculated as follows:

$$\mathrm{MI\ score} = \log_2 \frac{O_{11}}{\frac{R_1 \times C_1}{N}} = \log_2 \frac{O_{11} \times \mathrm{N}}{R_1 \times C_1}$$

The Log Dice score is calculated as follows:

$$\log \mathrm{Dice} = 14 + \log_2 \frac{2 \times O_{11}}{R_1 + C_1}$$

In these equations, O_{11} is the observed frequency of the collocation, N is the number of tokens in the corpus, R_1 is the frequency of the node in the corpus, and C_1 is the frequency of the collocation in the whole corpus.

The Delta P score is calculated as follows:

Delta P = p (outcome | cue = present) − p (outcome | cue = absent)

The Delta P is the probability of the outcome given the cue (P(O|C)) minus the prob- ability of the outcome in the absence of the cue (P(O|-C))

Appendix B: Ethics form

CONSENT FORM

Project Title: Cognitive processing of adjective-noun collocations in Turkish and English:
Name of Researchers: Dogus Can Öksüz
Email: d.oksuz@lancaster.ac.uk

Please tick each box

1. I confirm that I have read and understand the instruction sheet for the above study. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily	☐
2. I understand that my participation is voluntary and that I am free to withdraw at any time during my participation in this study and within 3 weeks after I took part in the study, without giving any reason. If I withdraw within 3 weeks of taking part in the study my data will be removed.	☐
3. I understand that any information given by me may be used in future reports, academic articles, publications or presentations by the researcher/s, but my personal information will not be included and I will not be identifiable.	☐
4. I understand that my name/my organisation's name will not appear in any reports, articles or presentation without my consent.	☐
5. I understand that data will be kept according to University guidelines for a minimum of 10 years after the end of the study.	☐
6. I agree to take part in the above study.	☐

________________________ _______________ ________________

Name of Participant Date Signature

I confirm that the participant was given an opportunity to ask questions about the study, and all the questions asked by the participant have been answered correctly and to the best of my ability. I confirm that the individual has not been coerced into giving consent, and the consent has been given freely and voluntarily.

Signature of Researcher /person taking the consent__________________________ **Date** __________ Day/month/year

One copy of this form will be given to the participant and the original kept in the files of the researcher at Lancaster University

Appendix C: Questionnaire

Biographical Questionnaire: It would be appreciated if you could take some time to complete this questionnaire.

Please answer the questions below.

1. Age: ______________
2. Gender:____________
3. Your dominant hand: ☐Right ☐ Left
4. Do you have normal vision with/without glasses or contact lenses?
   ☐ Yes ☐ No
5. What is your highest level of education completed?
   ☐ Postgraduate ☐ Undergraduate ☐ High school
6. Your department:________________
7. Have you ever been diagnosed with a learning difficulty?
   ☐ Yes ☐ No
8. Home language: ☐ English ☐ Turkish ☐ Other (__________)
9. How long have you been studying (English)? ______________
10. What languages other than English/Turkish do you know?
    ____
11. Have you ever taken an English proficiency exam (e.g. IELTS, TOEFL BUEPT)?
    ☐ Yes ☐ No
12. Which exam?___________________
13. When did you take it? (Please provide month and year)

______/_______

14. What were your scores?
    Listening:___

    Reading:__

    Writing:___

    Speaking:___

    Overall:___

15. Have you ever lived in an English speaking country?

☐ Yes   How long: _______________?   ☐ No

16. Rate your current overall language ability in English?

    1 = understand but cannot speak
    2 = understand and can speak with great difficulty
    3 = understand and speak but with some difficulty
    4 = understand and speak comfortably, with little difficulty
    5 = understand and speak comfortably
    6 = understand and speak fluently like a native speaker

17. On a scale from 1 to 6, rate your abilities in English (1=Beginner, 2=Low-intermediate, 3=Intermediate, 5=High intermediate 6=Advanced)
    Reading =   Speaking=   Listening=   Writing=

18. How much time **a day** (including class time) do you read in English or Turkish (Please tick the appropriate answer.)

	4 hours or less	5 to 8 hours	9 to 12 hours	13 to 16 hours	17 to 20 hours	more than that
English	☐	☐	☐	☐	☐	☐
Turkish	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐
	☐	☐	☐	☐	☐	☐

19. Is there anything else about this experiment you would like to comment on?

____

____

Appendix D: Lextale Placement test Items

0	platery	0
0	denial	1
0	generic	1
1	mensible	0
2	scornful	1
3	stoutly	1
4	ablaze	1
5	kermshaw	0
6	moonlit	1
7	lofty	1
8	hurricane	1
9	flaw	1
10	alberation	0
11	unkempt	1
12	breeding	1
13	festivity	1
14	screech	1
15	savoury	1
16	plaudate	0
17	shin	1
18	fluid	1

19	spaunch	0
20	allied	1
21	slain	1
22	recipient	1
23	exprate	0
24	eloquence	1
25	cleanliness	1
26	dispatch	1
27	rebondicate	0
28	ingenious	1
29	bewitch	1
30	skave	0
31	plaintively	1
32	kilp	0
33	interfate	0
34	hasty	1
35	lengthy	1
36	fray	1
37	crumper	0
38	upkeep	1
39	majestic	1

40	magrity	0
41	nourishment	1
42	abergy	0
43	proom	0
44	turmoil	1
45	carbohydrate	1
46	scholar	1
47	turtle	1
48	fellick	0
49	destription	0
50	cylinder	1
51	censorship	1
52	celestial	1
53	rascal	1
54	purrage	0
55	pulsh	0
56	muddy	1
57	quirty	0
58	pudour	0
59	listless	1
60	wrought	1

Instructions

This test consists of about 60 trials, in each of which you will see a string of letters. Your task is to decide whether this is an existing English word or not. If you think it is an existing English word, you click on "yes", and if you think it is not an existing English word, you click on "no". If you are sure that the word exists, even though you don't know its exact meaning, you may still respond "yes". But if you are not sure if it is an existing word, you should respond "no". In this experiment, we use British English rather than American English spelling. For example: "realise" instead of "realize"; "colour" instead of "color", and so on. Please don't let this confuse you. This experiment is not about detecting such subtle spelling differences anyway. You have as much time as you like for each decision. This part of the experiment will take about 5 minutes. If everything is clear, you can now start the experiment."

Appendix E: English stimuli

High-frequency Collocations	Adjective frequency counts	Noun Frequency counts	Collocation Frequency Counts	Log Dice scores	MI-scores
Long time	5.69	6.13	4.61	7.28	3.15
Young people	5.45	6.03	4.53	7.36	4
Recent years	5.14	5.89	4.4	7.39	5.05
Local government	5.61	5.74	4.48	8.06	4.29
Old man	5.67	5.71	4.34	7.65	3.72
Labour party	5.38	5.54	4.59	9.1	6.09
Hot water	4.89	5.48	3.97	7.34	5.86
Left hand	5.6	5.47	3.92	7	3.34
Economic development	5.31	5.45	3.98	7.36	4.57
Nuclear power	4.85	5.44	4.08	7.81	6.44
Blue eyes	4.95	5.38	3.9	7.4	5.76
Common law	5.24	5.37	4.01	7.73	5.18
Main road	5.34	5.37	3.82	7.07	4.23
High level	5.53	5.37	3.97	7.47	4.11
Social policy	5.56	5.36	3.89	7.19	3.74
Mental health	4.7	5.33	3.93	7.69	6.82
Prime minister	5.03	5.32	4.93	10.95	9.09
Annual report	4.85	5.32	3.75	7.13	5.78
Front door	5.27	5.31	4.16	8.4	5.8
European community	5.25	5.31	4.26	8.75	6.2
Free trade	5.26	5.24	3.71	7.08	4.54
Middle class	5.05	5.2	3.92	8	6.1
White paper	5.31	5.17	4	8.11	5.57
Bad news	5.12	5.1	3.78	7.8	5.72
Dark hair	5.05	5.08	3.49	7	5.09
Human rights	5.23	5.05	4.08	8.87	6.51
Rapid growth	4.5	5.05	3.47	7.09	6.9
Medical treatment	4.91	5.03	3.61	7.53	6.08
Vast majority	4.63	4.94	3.88	8.79	8.21
Senior officer	4.86	4.9	3.53	7.61	6.4

Low-frequency collocations	Adjective frequency counts	Noun Frequency counts	Collocation Frequency Counts	Log Dice scores	MI-scores
Inner world	4.59	5.7	2.85	2.92	2.39
Difficult life	5.28	5.69	3.11	3.76	1.01
Lovely house	4.73	5.64	2.79	2.95	1.97
Warm place	4.77	5.62	2.86	3.19	2.06
Poor children	5.11	5.6	3.63	3.63	1.43
Similar case	5.21	5.6	2.82	3.1	0.61
Whole company	5.42	5.54	2.76	3.01	-0.13
Certain point	5.28	5.54	3	3.84	1.09
Vital information	4.65	5.53	2.98	3.91	3.2
Small head	5.58	5.5	2.81	3.21	-0.36
Foreign business	5.15	5.49	2.68	2.84	0.65
Round face	5.43	5.46	2.85	3.54	0.4
Great service	5.61	5.43	2.78	3.3	-0.08
Tiny room	4.66	5.4	2.75	3.55	2.82
Special court	5.28	5.4	2.64	3.09	0.4
Easy question	5.1	5.35	2.49	2.76	0.64
Elderly mother	4.64	5.33	2.41	2.7	2.04
Physical body	4.92	5.34	2.74	3.69	2.13
Strong voice	5.14	5.34	2.67	3.47	1.24
Important city	5.53	5.3	2.57	3	-0.32
Fair idea	4.9	5.27	2.43	2.9	1.43
Good land	5.85	5.27	2.79	3.49	-0.53
Extra hours	4.92	5.21	2.65	3.8	2.28
Full authority	5.4	5.21	2.65	3.6	0.7
Only friends	6.12	5.16	2.66	2.78	-1.5
New award	6.04	5.12	2.93	3.85	-0.26
Away game	5.62	5.11	2.32	2.54	-0.81
Soft material	4.71	5.07	2.36	3.31	2.45
Suitable software	4.72	4.92	2.04	2.77	1.92
Dry glass	4.75	4.92	2.07	2.87	1.94

Baseline items	Adjective frequency counts	Noun Frequency counts	Collocation Frequency Counts	Log Dice scores	MI-scores
Dirty time	4.37	6.13	1.41	-3.22	-3.03
Sudden people	4.56	6.03	1.54	-2.48	-2.92
Nice years	5.05	5.89	1.54	-2.05	-4.1
Deep government	4.95	5.74	1.23	-2.53	-4.26
Far man	5.54	5.71	1.9	-0.41	-3.97
Hard party	5.29	5.54	0	0	0
Late water	5.23	5.48	1.69	-0.19	-2.76
Current hand	5.1	5.47	1.64	-0.39	-2.53
Able development	5.42	5.45	0.94	-2.21	-2.75
Sure power	5.32	5.44	1.54	-0.07	-3.5
General eyes	5.53	5.38	0	0	0
Regular law	4.82	5.37	1.41	-0.76	-2.01
Basic road	4.99	5.37	1.23	-1.38	-3.15
Big level	5.34	5.37	1.54	-0.48	-3.32
Huge policy	4.83	5.36	1.23	-1.3	-2.58
Single health	5.2	5.33	1.23	-1.32	-3.75
Real minister	5.3	5.32	1.23	-1.34	-4.04
Red report	5.11	5.32	1.54	-0.25	-2.38
Political door	5.42	5.31	1.23	-1.36	-4.4
Short community	5.24	5.31	1.54	-0.26	-2.77
Clear trade	5.34	5.24	1.69	0.45	-2.32
Public class	5.53	5.2	1.64	0.16	-3.07
Outside paper	5.27	5.17	0	0	0
True news	5.19	5.1	1.41	-0.059	-2.34
British hair	5.49	5.08	1.23	-0.83	-3.28
Low rights	5.17	5.05	0	0	0
Chief growth	4.99	5.05	0	0	0
Wide treatment	5.02	5.03	1.23	-0.4	-2.12
Light majority	5.3	4.94	0.94	-1.31	-3.75
Total officer	5.19	4.9	0.94	-1.11	-3.24

Baseline items	Adjective frequency counts	Noun Frequency counts	Collocation Frequency Counts	Log Dice scores	MI-scores
Necessary world	5.2	5.7	1.64	-1.14	-3.64
Final life	5.13	5.69	1.9	-0.22	-2.51
Fast house	4.81	5.64	1.23	-2.2	-3.45
Firm place	5.03	5.62	1.64	-0.86	-2.81
Official children	4.93	5.6	1.23	-2.11	-3.7
Dead case	5.02	5.6	1.41	-1.52	-3.7
Green company	5.1	5.54	1.72	-0.36	-2.51
Top point	5.35	5.54	1.23	-2	-4.94
United information	5.23	5.53	1.54	-0.92	-3.5
Various head	5.13	5.5	1.79	-0.03	-2.24
Married business	4.93	5.49	0.94	-2.74	-2.74
National face	5.53	5.46	1.79	-0.028	-3.41
Natural service	5.09	5.43	1.41	-0.99	-3.1
Key room	5.04	5.4	1.41	-0.91	-2.85
Happy court	5	5.4	1.23	-1.44	-3.28
French question	5.17	5.35	1.23	-1.38	-3.71
Significant mother	5.02	5.33	0	0	0
Early body	5.47	5.34	1.23	-1.46	-4.64
Sorry voice	4.98	5.34	1.23	-1.2	-2.93
Serious city	5.03	5.3	1.23	-1.2	-3.06
Cold idea	5.01	5.27	0.94	-2.12	-3.89
Ready land	4.94	5.27	1.41	-0.46	-2.07
Large hours	5.48	5.21	1.54	-0.13	-3.25
Future authority	5.29	5.21	1.23	-1.03	-3.62
Wrong friend	5.14	5.16	1.41	-0.2	-2.36
Obvious award	4.86	5.12	0	0	0
Individual game	5.22	5.11	0.94	-1.69	-4.06
Past material	5.35	5.07	1.23	-0.68	-3.36
Little software	5.74	4.92	0.94	-1.85	-5.15
Male glass	4.98	4.92	0	0	0

Appendix F: Turkish stimuli

High-frequency collocations	Adjective frequency counts	Noun Frequency counts	Collocation Frequency Counts	Log Dice scores	MI-scores
Genç adam	5.61	5.69	4.22	7.35	3.77
Sivil toplum	5.09	5.36	4.67	10.01	8.1
Ertesi gün	4.99	6.03	4.77	8.19	6.52
Milli egitim	5.45	5.7	4.61	8.66	5.56
Bati avrupa	5.32	5.57	4.12	7.49	4.83
Soguk savas	5.13	5.36	4.13	8.18	6.15
Yerel yönetim	5.23	5.38	4.16	8.21	5.9
Bilimsel arastirma	5.08	5.29	3.72	7.04	5.11
Uzun dönem	5.84	5.21	3.99	7.58	3.83
Dis ticaret	5.52	5.25	4.59	9.83	6.76
Erken seçim	5.11	5.18	3.64	7.12	5.23
Orta sinif	5.42	5.15	3.87	7.78	5.05
Kimyasal madde	4.76	5.16	3.56	7.01	6.18
Genel baskan	5.86	5.16	4.43	9.12	5.41
Yakin iliski	5.58	5.17	3.84	7	4.34
Yabanci sermaye	5.48	5.24	4.25	8.75	5.8
Sosyal güvenlik	5.67	5.29	4.65	9.82	6.35
Geçen yil	5.62	5.88	4.66	8.21	4.54
Kisa süre	5.62	5.65	4.81	9.41	5.83
Kücük kiz	5.85	5.54	4.25	7.7	3.56
Yogun bakim	5.21	4.52	3.46	8.03	6.48
Olumlu yanit	5.21	4.99	3.59	7.42	5.35
Agir ceza	5.42	5.03	3.89	8.11	5.47
Serbest meslek	5.17	5.03	3.57	7.27	5.26
Yüksek faiz	5.79	5	3.73	7.16	3.83
Kirmizi sarap	5.15	4.59	3.36	7.66	6.1

Low-frequency collocations	Adjective frequency counts	Noun Frequency counts	Collocation Frequency Counts	Log Dice scores	MI-scores
Güçlü kadin	5.33	5.79	3.16	3.6	0.86
Küresel dünya	4.87	5.78	3.07	3.38	2.11
Yalniz çocuk	5.46	5.67	3.49	3.57	0.37
Çabuk karar	4.96	5.58	3.06	3.98	2.38
Kötü durum	5.46	5.51	3.2	4.5	1.47
Aci haber	5.16	5.34	2.85	3.98	1.86
Siyasi hayat	5.45	5.35	2.75	3.52	0.58
Görsel sanat	4.51	5.31	2.51	3.1	3.03
Nitelikli isçi	4.53	5.25	2.69	3.87	3.75
Haksiz vergi	4.62	5.27	4.14	3.64	3.72
Sürekli görev	5.5	5.24	2.77	3.8	0.79
Askeri destek	5.31	5.2	2.78	4.09	1.6
Kalin kitap	4.79	5.3	2.69	3.66	2.7
Yanlis tercih	5.38	5.23	2.56	3.26	0.57
Ulusal hukuk	5.35	5.21	2.89	4.4	1.82
Belirli konu	5.23	5.33	2.74	3.61	1.31
Farkli düsünce	5.79	5.19	3.04	4.5	0.91
Kolay iletisim	5.4	5.2	2.75	3.94	1.22
Ilgili ögrenci	5.92	5.1	2.88	3.97	0.27
Fazla üretim	5.88	5.42	3.21	4.5	0.42
Güzel göz	5.84	5.49	2.97	3.59	-0.46
Benzer ülke	5.33	5.42	2.69	3.17	0.54
Tüm hafta	5.88	5.34	2.94	3.78	-0.21
Özel örnek	5.89	5.32	2.88	3.61	-0.37
Ana yemek	5.47	5.27	3	4.49	1.57
Ortak sonuç	5.43	5.29	2.74	3.64	0.84

Baseline items	Adjective frequency counts	Noun Frequency counts	Collocation Frequency Counts	Log Dice scores	MI-scores
Kuzey adam	5.09	5.69	0	0	0
Hizli toplum	5.31	5.36	0	0	0
Ileri gün	5.41	6.03	0	0	0
Kara egitim	5.22	5.7	0	0	0
Mümkün avrupa	5.56	5.57	0	0	0
Uygun savas	5.64	5.36	0	0	0
Hos yönetim	5.09	5.38	0	0	0
Canli arastirma	5.06	5.29	0	0	0
Ünlü dönem	5.21	5.21	0	0	0
Derin ticaret	5.23	5.25	0	0	0
Geçici seçim	4.89	5.18	0	0	0
Açik sinif	5.65	5.15	0	0	0
Yasli madde	5.18	5.16	0	0	0
Önemli baskan	6.07	5.16	0	0	0
Beyaz iliski	5.34	5.17	0	0	0
Yazili sermaye	4.99	5.24	0	0	0
Demokratik güvenlik	5.1	5.29	0	0	0
Gerçel yil	5.57	5.88	0	0	0
Hazir süre	5.15	5.65	0	0	0
Egemen kiz	4.93	5.54	0	0	0
Karanlik bakim	5	4.52	0	0	0
Kültürel yanit	5.2	4.99	0	0	0
Bos ceza	5.25	5.03	0	0	0
Emin meslek	5.01	5.03	0	0	0
Çagdas faiz	5.03	5	0	0	0.
Zor sarap	5.46	4.59	0	0	0

Baseline items	Adjective frequency counts	Noun Frequency counts	Collocation Frequency Counts	Log Dice scores	MI-scores
Temel kadin	5.66	5.79	0	0	0
Sicak dünya	5.21	5.78	0	0	0
Kesin çocuk	5.17	5.67	0	0	0
Bagli karar	5.58	5.58	0	0	0
Resmi durum	5.27	5.51	0	0	0
Temiz haber	4.87	5.34	0	0	0
Dogru hayat	5.96	5.35	0	0	0
Sert sanat	5.04	5.31	0	0	0
Büyük isçi	6.24	5.25	0	0	0
Mutlu vergi	5.24	5.27	0	0	0
Yavas görev	5.35	5.24	0	0	0
Rahat destek	5.14	5.2	0	0	0
Dogal kitap	5.42	5.3	0	0	0
Tam tercih	5.82	5.23	0	0	0
Genis hukuk	5.43	5.21	0	0	0
Yesil konu	5.11	5.33	0	0	0
Koca düsünce	4.98	5.19	0	0	0
Eski iletisim	5.74	5.2	0	0	0
Yasal ögrenci	5	5.1	0	0	0
Uzak üretim	5.27	5.42	0	0	0
Geleneksel göz	5.08	5.49	0	0	0
Hafif ülke	5	5.42	0	0	0
Basit hafta	5.08	5.34	0	0	0
Bütün örnek	6.08	5.32	0	0	0
Merkez yemek	5.2	5.27	0	0	0
Ince sonuç	5.22	5.29	0	0	0

Appendix G: Translation of the Turkish stimuli

High-frequency collocations	Low-frequency collocations	Baseline items	Baseline items
Young man	Strong woman	North man	Main woman
Civil society	Global world	Quick society	Hot world
Next day	Lonely child	Future day	Certain child
National education	Quick decision	Black education	Tied decision
Western Europe	Bad situation	Possible Europe	Official situation
Cold war	Sad news	Suitable war	Clean news
Local government	Political life	Nice government	Correct life
Scientific research	Visual art	Lively research	Tough art
Long term	Qualified worker	Famous term	Big worker
Foreign trade	Unfair taxation	Deep trade	Happy taxation
Snap election	Continous duty	Temporary election	Slow duty
Middle class	Millitary support	Open class	Comfortable support
Chemical substance	Thick book	Old substance	Natural book
General president	Wrong choice	Important president	Full choice
Close relationship	National law	White relationship	Wide law
Foreign capital	Selected topic	Written capital	Green topic
Social security	Different idea	Democratic security	Massive idea
Last year	Easy communication	Real year	Old communication
Short time	Careful student	Ready time	Legal student
Little girl	Excessive production	Soverign girl	Far production
Intensive care	Beatiful eye	Dark care	Traditional eye
Positive answer	Similar country	Cultiural answer	Light country
Heavy punishment	Full week	Empty punishment	Simple week
Self employment	Special example	Sure employment	Whole example
High interest	Main course	Contemporary interest	Central course
Red wine	Common result	Difficult wine	Thin result

Appendix H: The graphs for the Mixed-effect models

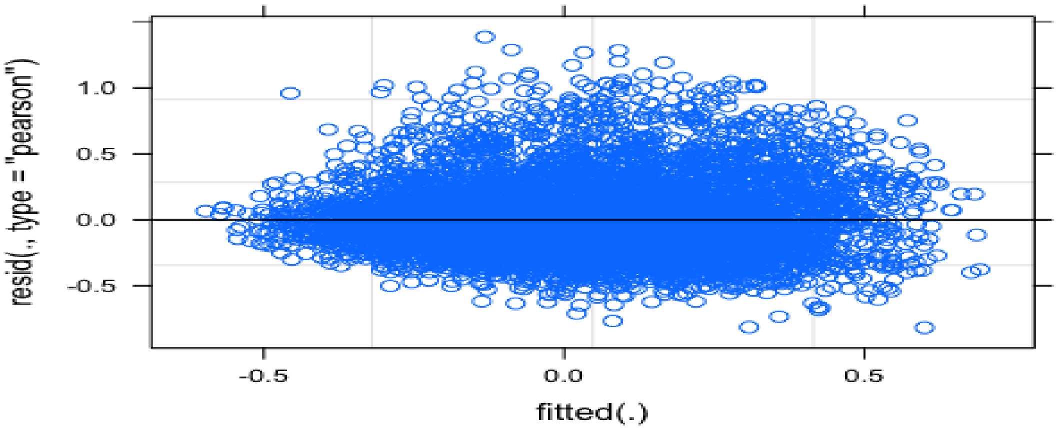

Figure H1: Residual versus fitted plot for Mixed Effect Model 1

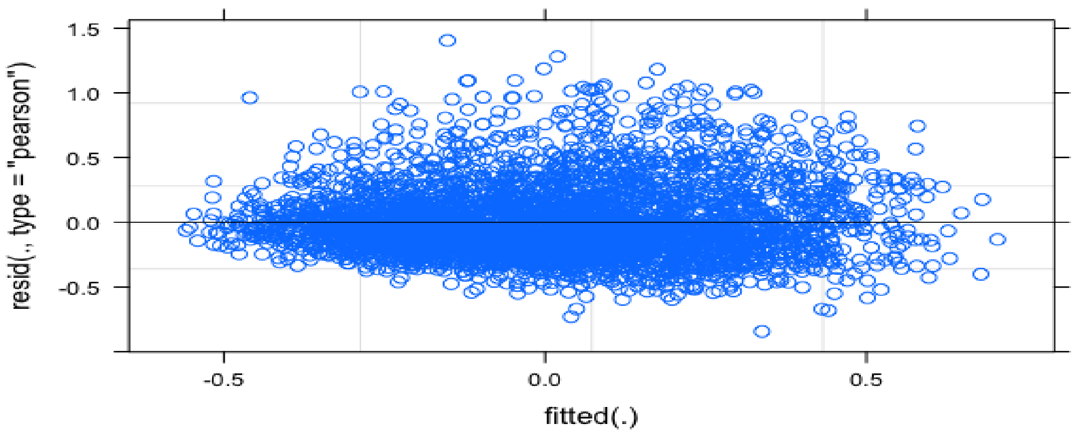

Figure H2: Residual versus fitted plot for Mixed Effect Model 2

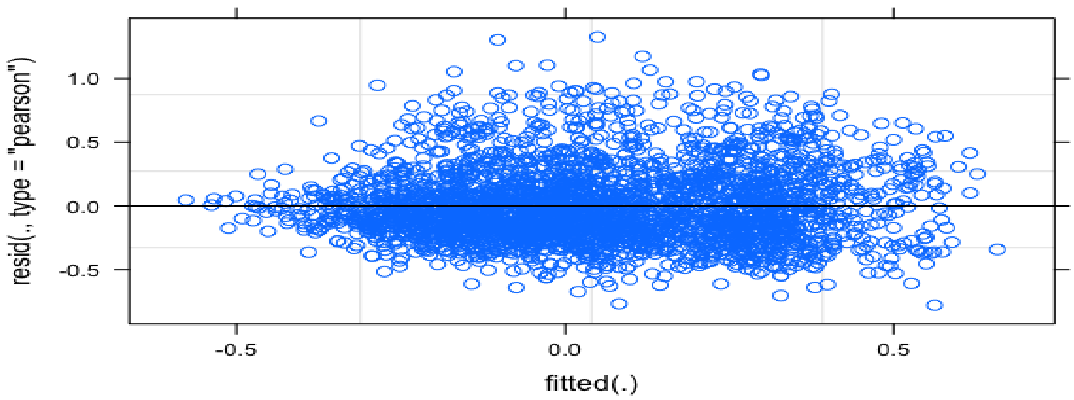

Figure H3: Residual versus fitted plot for Mixed Effect Model 3

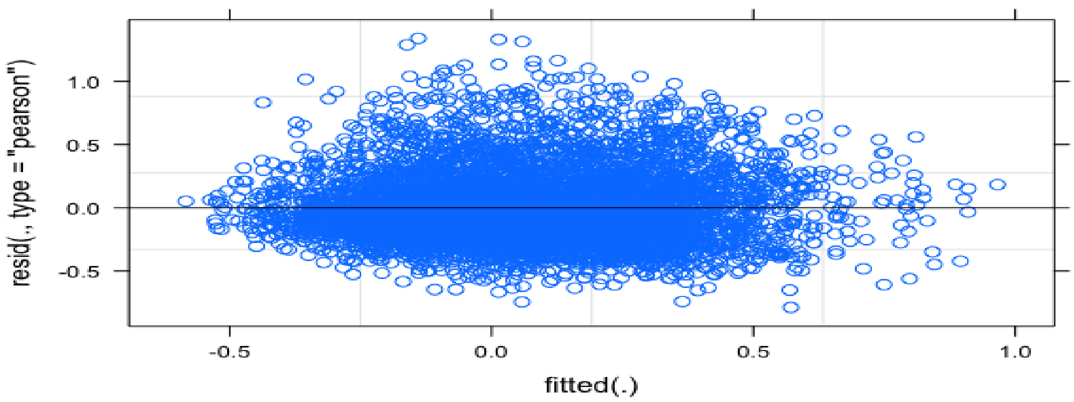

Figure H4: Residual versus fitted plot for Mixed Effect Model 4

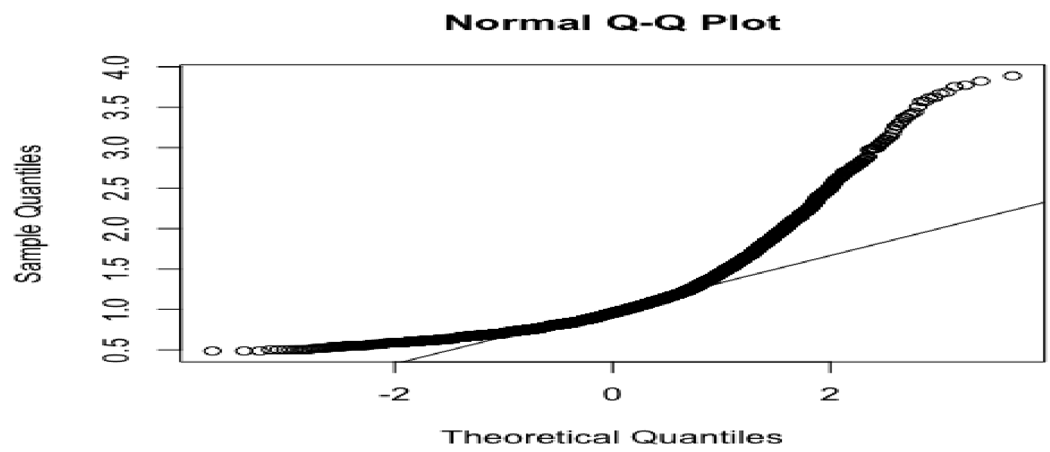

Figure H5: Q-Q plot for raw RTs in Turkish (Mixed Effect Model 2)

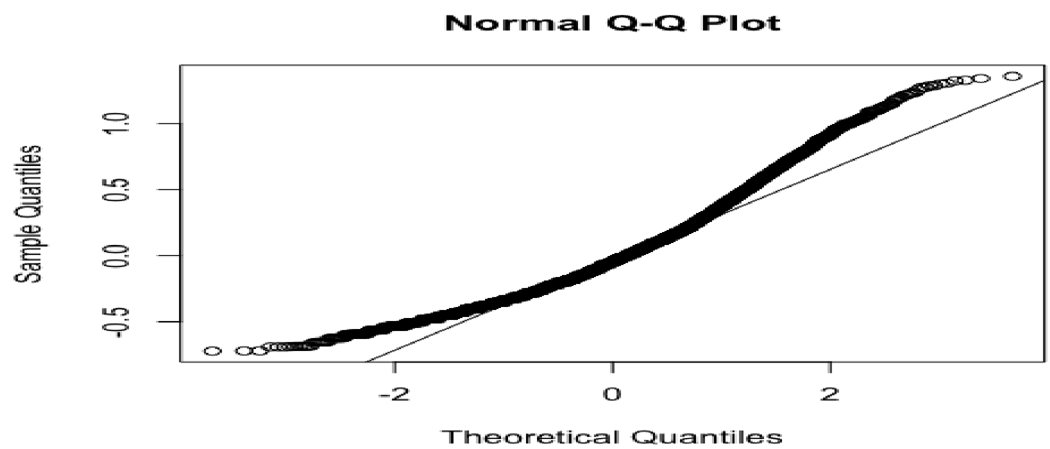

Figure H6: Q-Q for logged RTs in Turkish (Mixed Effect Model 2)