Clarence Green*

Usage-based linguistics and the magic number four

DOI 10.1515/cog-2015-0112
Received October 26, 2015; revised January 28, 2017; accepted February 13, 2017

Abstract: Miller's (1956, The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2). 81–97) working memory (WM) capacity of around seven items, plus or minus two, was never found by usage-based linguists to be a recurrent pattern in language. Thus, it has not figured prominently in cognitive models of grammar. Upon reflection, this is somewhat unusual, since WM has been considered a fundamental cognitive domain for information processing in psychology, so one might have reasonably expected properties such as capacity constraints to be reflected in language use and structures derived from use. This paper proposes that Miller's (1956) number has not been particularly productive in usage-based linguistics because it turns out to have been an overestimate. A revised WM capacity has now superseded it within cognitive science, a "magic number four plus or minus one" (Cowan 2001, The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24(1). 87–185). This paper suggests, drawing on evidence from spoken language corpora and multiple languages, that a range of linguistic structures and patterns align with this revised capacity estimate, unlike Miller's (1956), ranging from phrasal verbs, idioms, n-grams, the lengths of intonation units and some abstract grammatical properties of phrasal categories and clause structure.

Keywords: usage-based linguistics, working memory, corpus linguistics, cognitive grammar, psycholinguistics

1 Introduction

An in-principle stipulation of usage-based linguistics is that general cognitive domains should turn out to be sufficient to explain many of the properties of human language (Evans 2014; Evans and Levinson 2009; Langacker 2009). Usage-

***Corresponding author: Clarence Green,** Department of English Language and Literature, National Institute of Education, Nanyang Technological University, Singapore,
E-mail: clarence.green@nie.edu.sg

based linguistics with a cognitive orientation therefore involves exploring the interaction of the linguistic system with other cognitive processes (Givón 2002; Goldberg 2006; Lakoff 2009), and its progression requires continually integrating experimental and theoretical findings from the rest of cognitive science. This paper focuses on some relatively recent developments in working memory (henceforth WM) research, which is one of the most important cognitive domains for information processing (Henry 2012). WM has been an intensely studied subfield of cognitive psychology since at least mid-way through the previous century (Aboitiz et al. 2006; Baddeley 2003; Cowan et al. 2005). Arguably, the most influential paper in the history of WM research was Miller's (1956), in which he proposed that the capacity of WM was seven items, plus or minus two. Miller's (1956) famous phrase for this constraint was the *magic number seven*. The magic of the number, he suggested, derived from its recurrent patterning with psychometric tests, both visual and auditory. Miller (1956) also argued that WM operates in a hierarchical fashion and its capacity can be extended by chunking. That is, lower order sequences of items can form chunks, and these chunks can function as single units in a higher order sequence.

A cognitive domain that constrains information processing and operates hierarchically is highly relevant to usage-based linguistics. Bybee (2006, 2010) invokes the process of chunking to explain how sequences of words become lexicalized into single symbolic units over time, as in the archetypal example of *be going to*. Multi-word sequences such as formulaic language have also often been explained with reference to WM processes and chunking (Biber et al. 2004; Wray 2002). Even though the word has played a prominent role in WM research (Baddeley 2003), it is not of course the only information upon which WM operates. It has been proposed to limit the length of reference chains in discourse (Givón 1984; Gernsbacher and Givón 1995), constrain information packaging in the clause (Chafe 1994), play a role in the integration of discourse propositions (Gilchrist et al. 2008; Van Dijk and Kintsch 1983), and in general, an individual's working memory (typically measured by word/digit span recall) correlates highly with first and second language proficiency (Gathercole et al. 2003; Ellis and Sinclair 1996).

However, while chunking has been a productive concept in research, usage-based linguists never found that Miller's (1956) magic number seven was a recurrent property of language, and thus it has never been considered as a feature of cognition that might contribute to a model of cognitive grammar. Upon reflection this is somewhat odd, as it seems reasonable to expect that if WM is a cognitive domain that constrains language use, and structure derives from use, then a relationship would be expected. This paper argues that perhaps the reason why usage-based linguistics has not been able to make much use of

Miller's (1956) capacity estimate is because it was an overestimate. There has been a relatively recent correction beginning with the seminal article by Cowan (2001), which has revised WM capacity down to four items, plus or minus one. This new *magic number four*, has in general superseded Miller's (1956) larger estimate in cognitive science (Bor 2016), yet this revision could do with more exposure in linguistics, since Miller's (1956) number is still cited in otherwise up-to-date linguistic research (Mahlberg et al. 2014). As Van Dyke and Johns (2012: 193) note, "linguistic theory has generally not made close contact with memory research" and even though one finds the revised capacity of WM now cited in the most introductory of cognitive psychology textbooks (Hills and Pake 2016), there has been little discussion or research attention to this significant development from within the field of cognitive linguistics. This should be corrected as the field is in spirit and practice the linguistics branch of cognitive science. Thus, the following paper has two modest goals. First, it opens the discussion space for the revised WM capacity and its implications in the context of usage-based linguistics. Second, it offers preliminary evidence that unlike the magic number seven, the revised WM capacity of four, plus or minus one, appears to a recurrent linguistic pattern. The general contention is that WM constrains use and that language behavior demonstrates an alignment with the revised capacity of four, plus or minus one, manifesting in linguistic phenomena such as phrasal verbs, idioms, n-grams, phrase and clause structures, and intonation unit length. Data is drawn from a range of spoken corpora across multiple languages. As Givón (2012: 48) puts it so well, what follows may turn out to "be only partially accurate… but this kind of partial accuracy is a necessary early step in all science".

The paper is organized as follows. Section 2 outlines the current model of WM, providing an overview of some key concepts and terminology relevant to usage-based linguistics. Section 3 considers phrasal verbs, idioms and n-grams in relation to capacity constraints, and Section 4 does so for noun and verb phrases. Section 5 extends the analysis to clause grammar and Section 6 to intonation units. A general discussion is had in Section 7.

2 Working memory, chunking and language

The most prominent current model of WM is that of Baddeley (2003, 2007), in which WM is the intermediate cognitive domain between long term memory (LTM), where procedural and semantic memory is stored, and sensory memory, where the physical world first interacts with cognition (Henry 2012). Baddeley's

(2007) is not the only model of working memory in cognitive science but it "has had immense popularity and influence" (Van Dyke and Johns 2012: 195). Baddeley (2003: 190) proposes WM consists of three subsystems: a phonological loop, an episodic buffer and a visual-spatial sketchpad. These are "slave systems" controlled by the central executive, which directs attentional resources. The episodic buffer is a cognitive space where information from the phonological loop, visual-spatial sketchpad, and activated LTM can be integrated and symbolically represented (Robinson-Riegler and Robinson-Riegler 2008). The visual-spatial sketchpad, specialized for visual information, is often thought to be primary and to have evolved earlier than the phonological loop (Hurford 2007). However, it is the phonological loop that is mostly involved in language processing, and there have been several proposals that the phonological loop (co)evolved with language to facilitate language acquisition (Aboitiz et al. 2006; Baddeley 2003: 194). Speculatively, hierarchical combinatorially, fundamental properties of both WM and language, may derive from co-evolution. The phonological loop consists of two components (Baddeley 2003: 191): a phonological store and an articulatory loop. The phonological store holds incoming auditory information for a period of around 2 seconds (Baddeley 2007), while the articulatory loop is a mechanism for refreshing the phonological store by rehearsing information to keep it activated.

The phonological loop and the visual-spatial sketchpad are parallel subsystems that process visual and auditory information concurrently, and the capacity constraint of WM applies to each subsystem equally; that is, the visual-spatial sketchpad seems constrained to around four items at once, as with the phonological loop (Cowan et al. 2005). Cowan's (2001: 8) revision of WM capacity is based on a reevaluation of its "pure capacity limits". In his seminal article, he demonstrates through a review of decades of research that Miller's (1956) findings were based on experimental procedures that tapped into multiple processes. For example, Cowan (2001) shows that experimental measures based on recalling sequences of stimuli can lead to overestimates of WM capacity when participants are able to subitize some items (the instant formation of a temporary chunk) and subvocally rehearse other items. When these processes are teased apart in experiments that suppress rehearsal in the articulatory loop, then a consistent pure capacity limit emerges at around four items, plus or minus one; much lower than Miller's (1956).

Both Miller (1956) and Cowan (2001) agree, however, that WM extends its limited capacity via chunking. As noted above, chunking can consist of ephemeral associative bindings via subitization. However, recurrent associative bindings tend to become stored as chunks in LTM (Hurford 2007: 94). This process has been cited as a basic mechanism of language acquisition (Bannard and Lieven 2012; MacWhinney 2005). What counts as a chunk in WM operations has no simple definition. The definitions of *item, unit* and *chunk* in relation to WM are terms that are often broad and overlapping in the cognitive psychology literature. Depending on the level of analysis, the terms are interchangeable (Cowan et al. 2005: 60). For example, a word is an item at one level of analysis and has been the most common probe of WM since Miller (1956), yet it is also a chunk of syllables with high transitional probability at another level of processing (Ellis and Sinclair 1996). In the following paper, the term unit is used generally, with item referring to lower order units (often words), while chunk refers to higher order units (often multi-word sequences). Cowan's (2001: 89) definition of a chunk is "a collection of concepts that have close associations to one another". Note that "concepts" here may be misleading for linguists, since Cowan (2001) notes a chunk can be patterns of dots in an image or digit sequences, thus encompassing both information processing and symbolic representation. The definition is perhaps intentionally broad in order to capture the range of what is cited as a chunk in cognitive science. This has included subitized groupings in visual arrays (Hurford 2007), ad-hoc groupings of digits to facilitate the recall of numerical sequences (e. g., phone numbers) (Baddeley 2003), associative learning that groups new information with long term memory stores (Ericcson et al. 1980), and more besides. Neuroscientist Gazzinga (2015: 5) even cites the lexical item *truck* as a chunk, on the basis that it is a neural binding of its category of vehicle, its function to carry things, and its component parts of having an engine etc. What counts as a chunk therefore seems to be a continuum, from units that may be neurally bound and fixed, through to units made up of items with high transitional probability, and even only temporary cognitive associations. In linguistics, the term has also had wide application, including phonemes chunked into syllables (Kuhl 2004), syllables into words (Gathercole and Alloway 2008), words into idioms and formulaic language (Sosa and MacFarlane 2002), words into lexicalized phrases (Bybee 2010), concepts within propositions (Hurford 2007), words bound into phrasal categories (Bergen 2012), and intonation units within discourse (Chafe 1994).

Though chunking has often been invoked by linguists with the understanding that it is a mechanism of WM to overcome capacity constraints, the revised capacity constraint on WM has not received the same attention. This is unfortunate since as early Cowan's (2001) target article, Pothos and Juola (2001:138) argued in a direct response that the revised pure capacity limit is reflected across multiple languages. They begin with the proposition, which is essentially the same as in this paper, that one would expect language to reflect WM capacity if language is shaped by use and there are constraints on use. They calculated the entropy of word sequences in eight languages represented by the European Corpus Initiative Multilingual Corpus and computed Mutual Information (MI) scores across all word sequences. What they found was that at around four words in all of the languages, MI scores declined toward zero, indicating that words tended to have statistically significant relationships with other words up to around four words, after which there were few to no correlations. They conclude that this reflects the influence on language of the WM capacity proposed by Cowan (2001). Falsifiability is achieved in this corpus study in the sense that MI scores spanning closer to 7 word sequences would have been consistent with Miller's (1956) larger estimate, while MI scores insignificant after fewer words than four might have suggested no relationship between WM capacity and multi-word sequences, or perhaps a smaller capacity.

The current study extends Pothos and Juola (2001) approach, and illustrates multiple other language patterns consistent with the revised WM capacity constraint. Like their work, some of these patterns relate to word sequences in spoken corpora, which is worth comment at the outset. Van Dyke and Johns (2012: 195) in review of WM capacity theories note that the word has a long history of being a "relevant metric" in cognitive psychology, but it not the only one and it is an imperfect measure. In the following paper, it is not suggested that words in speech always map 1–1 to an item in WM attention, nor that all words have equivalent cognitive load during discourse processing (consider pronouns versus first mention lexemes, priming, common ground etc.) The word is but a proxy metric for measuring WM span, as in the corpus research of Pothos and Juola's (2001) and experimental research (see Cowan et al. 2008 for a review).

3 Multi-word units: Phrasal verbs, idioms and n-grams

English and perhaps most languages have many recurrent multi-word sequences (Bannard and Lieven 2012). These may be highly fixed sequences such as phrasal verbs, which are difficult to disambiguate from single lexical items (MacWhinney 2005), they may be formulas and idioms (Liu 2003; Wray 2002: 34), or they may simply be n-grams, i. e., high frequency recurrent phrases (Arnon and Snider 2010; Ibbotson and Tomasello 2009). The terminology for such sequences, let alone their cognitive status, is far from settled across the literature. A reasonable view is that a gradient exists along which phrasal verbs, idioms, formulaic sequences and n-grams distribute. As Xiao (2010: 4) states "Idioms ... are similar to fixed and semi-fixed formulaic expressions based on collocations which are known as word clusters, lexical bundles, multiword units, prefabs, and n-grams … the demarcation line between idioms and word clusters is actually fuzzy … [Idioms] can be regarded as an extreme example of word clusters". In general, however, idioms and phrasal verbs consist of relatively fixed word sequences that seem to be produced and processed as single units in discourse (MacWhinney 2005; Sinclair 1991). Part of the evidence that has been offered for their status as chunks is that their semantic profile spreads across the words in the unit (Martinez and Schmitt 2012: 300), and they are processed faster in experimental settings in comparison to word sequences with lower transitional probabilities (Sosa and MacFarlane 2002). N-grams are a broader linguistic phenomenon, within which idioms and phrasal verbs are special cases. N-grams are high frequency recurrent word strings that may be fully compositional and cross phrase category boundaries. Again, their cognitive status is debated (Siyanova-Chanturia 2015); however, what has been demonstrated experimentally is that phrase frequency alone has cognitive advantages (Bannard and Matthews 2008; Sosa and MacFarlane 2002), and so current research suggests statistical information that spans recurrent sequences of words is stored cognitively (Arnon and Snider 2010).

These multi-word sequences all fall within the scope of Cowan's (2001) definition of a chunk as a collection of items with a close association to each other. The claim that WM allocates attentional resources to words during communication is a long-standing one in cognitive psychology (Baddeley 2007; Miller 1956; Van Dyke and Johns 2012) and it seems not unreasonable to examine large collections of authentic spoken language corpora to explore whether Cowan's (2001) revised capacity span might be a soft constraint limiting the lengths and distributions of multi-word sequences.

3.1 Corpus consultation: Idioms and phrasal verbs

To explore this, two corpora were created from the online dictionary WIKTIONARY. One contained all headword entries for English phrasal verbs and the other all entries for idioms. There is no guarantee that all English phrasal verbs and idioms are listed in the dictionary, and all entries are crowd sourced rather than selected by lexicographers or linguists (Liu 2003). However, web corpora have the advantage of size, and WIKTIONARY provided a sample (as of June 2016) of 2946 phrasal verbs and 8680 idioms. Table 1 reports the frequencies of phrasal verbs and idioms according to words per unit, with some data examples appended below.

Table 1: Frequencies of English phrasal verbs and idioms by length.

Phrasal Verbs					
2 words	3 words	4 words	5 words		
2683 (91 %)	237 (8 %)	25 (1 %)	1 (0.03 %)		
Idioms					
2 words	3 words	4 words	5 words	6 words	> 7 words
3701 (44 %)	2357 (28 %)	1402 (17 %)	488 (6 %)	221 (3 %)	163 (2 %)
N words	**Phrasal Verb**	**Idiom**			
2	*take up*	*big shot*			
3	*put up with*	*part and parcel*			
4	*come to terms with*	*keep an eye on*			
5	*fly in the face of*	*turn over a new leaf*			
6		*in any way, shape or form*			
7		*your guess is as good as mine*			

Observe that the number of phrasal verbs and idioms in the dictionary decline as the number of words in the unit increase. While 25 phrasal verbs were listed in the dictionary with four words, and one even with five, inspection of the data revealed all contained nominals, and were in fact three word phrasal verbs in a citation form, e. g., *get ahead of oneself*. The five-word sequence was *fly in the face of*, and it is debatable if this belongs in a phrasal verb dictionary. Regardless, note how this 5–word example has only two stress beats and three function words (all within the ten most frequent words of English). High resting activation and low cognitive load correlates with high frequency function words (Levelt et al. 1999). That the maximum length of English phrasal verbs is three words is no new finding, but its consistency with revised WM capacity has not been pointed out. Similarly, Table 1 shows that in almost nine thousand idioms, 89 % were four words or fewer. A reasonable linking hypothesis on the basis of the data is as follows: Bybee's (2006, 2010) "the mind's response to repetition" in which recurrent language patterns are chunked over time interacts with Cowan's (2001) revised WM capacity which places a soft constraint on the length of lexicalized multi-word sequences (i. e., chunks). Of course, not all of the data are lexicalized associative bindings of words. Behind these numbers are multi-word sequences acquired holistically during language acquisition via chunk extraction (Ibbotson and Tomasello 2009), and others such as proverbs are likely remembered wholesale via mechanisms similar to memorizing lines from a play rather than working memory associative binding. The longest idiom, which was 11 words, is likely such as case: *look as if one has lost a shilling and found sixpence*. While boundaries are fuzzy, Sinclair (1991) argued that idiomatic language is best restricted to those formulas that can seamlessly be slotted into otherwise fully creative utterances, and have comparatively high frequency in a corpus, and it is these that fall within the scope of our hypothesis.

The distributions in the dictionaries are skewed to within WM capacity, but there were several hundred multi-word idioms greater than four words. How these are used in speech was investigated, with the rationale that units greater than four, plus or minus one, would likely be rare given WM constraints. The WIKTIONARY data were therefore searched in the BNC spoken component (10 million words BrE). First, a wordlist index was created using WORDSMITH (Scott 2015), which was then clustered to produce recurrent sequences ranging from two to seven words, and occurring at a frequency greater than five (the upper limit of seven words was set to reflect Miller's (1956) original WM capacity). These BNC lists were imported into Excel and a function was written to match them against the dictionary data, providing the frequency estimates in Table 2.

Table 2: Frequencies of English phrasal verbs and idioms (BNC Spoken).

Phrasal Verbs	**2 words**	**3 words**	**4 words**	**5 words**	**6 words**	**7 words**
Types	947	89	7	0	0	0
Tokens	97501	4161	122	0	0	0
Proportion	95 %	4 %	1 %	0	0	0
Idioms	**2 words**	**3 words**	**4 words**	**5 words**	**6 words**	**7 words**
Types	1006	267	48	13	3	0
Tokens	276322	32328	3385	190	344	0
Proportion	88 %	10 %	1 %	0.06 %	0.1 %	0

Table 2 indicates that idioms listed in the dictionary longer than four words are not particularly productive in spoken language. While 28 % were four words or more in Table 1, actual use was 1.16 %. The observational evidence thus aligns with WM capacity. Note, however, that six word idioms had more tokens than five word idioms. This was due to 322 uses of, *at the end of the day*. It is certainly a circular argument to propose that this is not a chunk because it is six words and thus does not fit with the four, plus or minus one, model. Readers are of course welcome to consider this falsification, yet the argument is not that there is some hard boundary at four orthographic English words. It need not always be the word that is the salient information unit attended to (see section 5 and 6). Also consider that *at the end of the day* is a sequence of monosyllabic, mostly semantically empty high frequency words which have easy lexical access and low processing cost (Levelt et al. 1999). In terms of phonological structure, often the determiner forms a single phonological word with its noun, and the sequence above is single prosodic unit with only two heavy beats (Jackendoff 2003: 656). Certainly longer sequences can be chunked when they have features such as low complexity, low information value across words, incorporate smaller chunks, and form coherent prosodic units. Also, WM is not the only explanatory variable for the multi-word sequences, given the range of other factors that contribute to the formation of entrenched symbolic units well-known to cognitive linguists, e. g., metaphor, rhythm, metonymy, common ground etc. (Colston 2008; Langacker 2009).

Returning to the data, since two word phrasal verbs were the overwhelming majority in the dictionary, it is not surprising that they made up 95 % of all phrasal verbs in the spoken corpus. Perhaps smaller phrasal verbs are reduced because they are more frequently used (Haspelmath 2008), though it is unclear whether a Zipf effect would necessarily apply to multi-word units. Again, the four word phrasal verbs were of the kind *take it out on*, and questionable category members. One might ask, however, why there are not more three (and four) word phrasal verbs, given that this is within WM capacity. One consideration is that, from a diachronic perspective, some phrasal verbs that lexicalized from recurrent word sequences emerged within VPs with slots for tense, aspect, and mood. For example, Denison (2004: 153) notes that *put up with* emerged through a reanalysis of three phrasal categories to one: *He [put] [up [with X] > he [put up with] X*. During lexicalization, usage would have included *have/has/could have put up with* and so forth. As discussed in Section 4, WM might also constrain the length of VPs and this may limit the length of multi-word verbs that can emerge.

Let us consider some cross-linguistic idioms, since WM capacity constraints and operations such as chunking should not be only reflected in English (Hurford 2007). Mandarin is a highly analytic language, in which there is a much stronger correlation between the word, morpheme and syllable than in English. Generally, transliterations of English sentences require more words in Mandarin. Xiao (2010) explored the distribution of idioms in multiple Mandarin corpora and offers a discussion of a class of idioms known as Chengyu. These are idioms that largely derive from classical Chinese sources, and by some estimates there are up to 20,000 thousand of them. An example is given in (1).

(1) Mandarin Idioms: Chenyu (adapted from Wu 1995)
唯 利 是 图
wei2 li4 shi4 tu2
ADV profit AUX work
Only working for profit

Chengyu idioms are formulaic patterns of four. Wu (1995: 67) argues that their influence has been so great “it might be said that most Chinese spoken expressions and written sentences continue to be based on the idiom’s four character form”. They are not the only idioms in Mandarin, and there are some that are longer and shorter. Xiao (2010) reports 7979 idioms in the Lancaster Corpus of Chinese (LCC), but does not report how many of these were Chengyu. His data was therefore rerun. All idioms in the Lancaster Corpus of Chinese are tagged <POS = “l”>, so all hits with this tag were extracted using WORDSMITH. Syllable lengths were computed by keying off tone unit transcriptions in the pinyin version of the corpus. Results are reported in Table 3 (n.b. the publically available LCC is smaller than the original study).

Table 3: Idioms by frequency and syllable length in the Lancaster Corpus of Chinese.

1 syllables	2 syllables	3 syllables	4 syllables	5 syllables	6 syllables	>7 syllables
2	116	697	3383	53	18	2

The average length of idioms in the Mandarin corpus was 3.79 syllables, and as Table 4 shows, the majority were Chengyu. One might suggest that this preference for patterns of four in idiomatic expressions in Mandarin reflects the general cognitive comfort of WM with four, plus or minus one.

Table 4: N-grams in the BNC spoken component (>40 p/m).

	2 words	3 words	4 words	5 words	6 words	7 words
Types	2645	523	34	2	0	0
Tokens	3570790	397263	19691	1115	0	0
Proportion	89.5 %	9.9%	0.5%	0.03 %	0	0

3.2 Corpus consultation: N-grams in bnc spoken

Idioms and phrasal verbs are special cases of n-grams. N-grams are any recurrent word sequence, as the examples in (2) illustrate.

(2) N-grams from BNC spoken
of the /very much /a lot of /to deal with /do you want to

In principle, an n-gram can be of any length, as they are extracted from data on the basis of the statistical co-occurrence. The emergence of n-grams as an object of study has largely derived from corpus linguistics, and while the psycholinguistic status of an n-gram (i. e., whether stored holistically or not) may need to be determined experimentally on a case-by-case basis (Schmitt et al. 2004; Siyanova-Chanturia 2015), what has been demonstrated experimentally is that phrase frequency alone, much like word frequency, is cognitively retained and has processing advantages (Arnon and Snider 2010; Sosa and MacFarlane 2002). If the span of WM attention is around four units, the statistical information that can be stored for recurrent word sequences might align with this.

From the BNC spoken all n-grams from two to seven words were computed, with a minimum frequency of forty per million words. This has been argued to be a reasonable lower boundary to isolate n-grams with psycholinguistic processing advantages (Biber et al. 2004). No other limitations were imposed on the WORDSMITH tool. Results are reported in Table 4, with proportions calculated against the total number of tokens.

The results indicate an n-gram distribution that aligns with Cowan's (2001) revised WM capacity, in that with spoken language, recurrent phrases of four or five words are less than 1% of all tokens, and none were detected beyond this. There is no long tail to the distribution; rather, n-grams approximate zero around the upper boundary of WM capacity. There do not seem to be any that align with Miller's (1956) WM capacity estimate of around 7. Not every n-gram Table 4 is a WM chunk, of course, and as Schmitt et al. (2004) emphasize experimentation is required to confirm which corpus-derived clusters have psycholinguistic processing advantages. Nevertheless, this data is essentially a different expression of Pothos and Juola's (2001) corpus finding that mutual information scores in numerous European languages become statistically insignificant after four words, which they cited as supporting evidence for the revised WM capacity.

4 Constraints on English NPs and VPs

Abstract grammatical categories, from a usage-based perspective, are generalizations across recurrent patterns in input (Hilpert 2014; Pothos 2007). The phrasal category is a grammatical relation that is hierarchical, and functions to bind words into symbolic units (Bergen 2012). Again, given that language use is constrained by WM which is itself a hierarchical cognitive process, one might expect grammar would reflect the properties of this cognitive domain. Typically the grammar of noun and verb phrases is described as a head with optional modifiers. The number of possible modifiers, and consequently the length, of these phrasal constituents is not restricted by the grammar. An NP head, in principle, accepts pre and post modification by any number of adjectives, adverbs, other nouns etc. Similarly, there is no limit to the number of adverbs permitted to modify a VP. Despite the principle of unbounded modification, there are in fact patterns of four that restrict English noun and verb phrases. Quirk et al. (1985), for example, analyze English NPs as constrained to a maximum string of three determiners before the head. These are pre-determiners, central determiners, and post-determiners. The NP in (3) illustrates this maximum string: *all* is a pre-determiner, *the* is a central determiner and *many* is a post-determiner.

(3) Maximal determiner modification in an NP (BNC spoken: K76)
NP > D D D N
all the many tasks

These are mutually exclusive grammatical classes that do not permit recursion. Once one of these determiner slots is filled, the grammar does not allow another of the same type in the string. Thus, NP grammar exhibits a grammaticalized pattern of four. Let us turn to the verb phrase. Even with the stipulation that unbounded adverbial modification is permitted, the English VP is restricted to four, potentially five, verbs, as in (4).

(4) Constraints on the VP

N words	Verb Phrase
1	*do*
2	*have done*
3	*has been done*
4	*could have been done*
5	*?could have been being done*

The only sequence in (4) not attested in the BNC (spoken) was the five word sequence: the modal passive continuous. It is questionable whether it is productive in spoken English. In sum, the phrasal categories of English exhibit certain grammatical properties consistent with the WM number cited by Cowan (2001). If grammar is an abstraction that emerges from recurrent encounters with input (Bybee 2006; Goldberg 2006), then it is not unreasonable to suggest the restrictions on the VP and NP grammatical categories might have a cognitive explanation, namely the (revised) WM capacity constraints.

4.1 Corpus consultation: NP modification

It has been suggested that the constraints on determiner strings in the NP may have its origins in use over time interacting with the WM cognitive domain. Yet, we have also acknowledged unbounded modification. If corpus data were to indicate that NPs are, in fact, frequently long, this might be taken as evidence against the proposal that an abstract grammatical pattern could derive from recurrent patterns of four. The BNC spoken was again examined to determine the frequencies of NPs ranging in length from two to eight words (head plus 1–7 modifiers). Data were extracted using a Perl script that searched for all substantive nouns and counted immediately preceding adjectives, determiners, adverbs and/or nouns up to a sentence break or verb. The script was somewhat limited, and was not able to count coordinated modification, e. g., *old and yellow potatoes*, coordinated NPs, e. g., *old potatoes and carrots*, nor gerund headed NPs. Results are reported in Table 5.

Table 5: Frequencies of noun phrases by length (BNC spoken).

Length	2 words	3 words	4 words	5 words	6 words	7 words	8 words
Freq.	588 210	177388	32486	4646	667	103	42
%	73 %	22 %	4 %	0.6 %	0.001	0.0001	0.00005

The data indicate that noun phrase length exhibits a similar pattern to n-grams and idioms in that relative frequency falls below 1 % after four words. Though there were several hundred tokens in the tail of the distribution, a qualitative check of the data indicated few convincing examples that were not artefacts of the search procedure. One of the longest modification sequences was: *a a single non-symmetrical a single hand gesture* [BNC: JSA]. This is a sequence of false starts, with multiple intonation contours, rather than an eight word NP. Also, the extraction script overestimated: e. g., *full well perhaps too much too much discussion, consultation, assessment* [BNC: HVG].

5 Clause level constraints

In terms of clause structure, English contains no verbs with a valency that licenses more than three arguments. Verbs such as *put* or *give* represent the maximum complementation pattern. The equivalent of *give* is the most common three argument verb cross-linguistically (Haspelmath 2003). The set of canonical clauses for English are given in (5) (Collins and Hollo 2009; Downing 2015; Huddleston and Pullum 2002).

(5) Canonical clauses of English

SV	*She sang* [BNC: PS1]
SVOd	*She changed bedrooms* [BNC: PSO]
SVC	*She felt sorry* [BNC: D91]
SVOiOd	*She gave me a key* [BNC: KCN]
SVOdOc	*She keeps him prisoner* [BNC: GV8]

The canonical clauses of English reflect a limit on the number of phrase categories in a clause that have necessary grammatical relations to each other. What is meant by a necessary grammatical relation is that an object complement (Oc) depends on an object (Od), which depends on a transitive verb (V) and a subject (S). Thus, SVOiOd, for example, forms a string in which each phrasal category is grammatically coded as a core argument. While the clauses in (5) may add any number of adjuncts, these phrasal constituents are not coded as core arguments (excepting SVA and SVOA clauses with obligatory adjuncts such as *she put the book on the table*). In usage-based linguistics, clause patterns are essentially constructional schemas that emerge as abstractions over input (Goldberg 2006; Givón 2002). If WM is able to attend to only around four bindings of words into groups (i. e., phrasal categories) at a time, it is plausible that grammatical relations do not tend to emerge that extend beyond this.

5.1 Corpus consultation: Clause patterns

If the grammatical restrictions on clause structure have emerged via interaction with the attentional constraints of WM, one might expect that clauses are generally not greater than four phrasal constituents in spoken language even when all adjuncts are taken into account. Clause, phrase and function tags were searched in the Christine Treebank (Sampson et al. 2000), a 100 000 word parsed subsample of BNC spoken. Table 6 reports the distributions of canonical clauses, phrasal categories, and the total number of clauses.

Table 6: Clause and phrase distributions in the Christine Treebank.

SV	SVO	SVC	SVOI	SVOC	Adjuncts
2850	5423	2969	364	182	7906

Total phrasal categories	Total canonical clauses	Total clauses
45850	11788	15560

The data indicate that the majority of clauses were either SVO or SVC, and that the average number of phrasal categories per clause was 2.95. As a rough estimate, if the frequency of adjuncts is distributed across the total number of clauses, there is approximately one adjunct for every two clauses (51 %). In other words, clauses in spoken English do seem to be, on average, around three to four phrasal constituents.[1]

6 Intonation unit constraints

An *intonation unit* (IU) consists of words grouped under a single intonation contour, which are offset from the ongoing discourse by a slight pause. Interestingly, there has been a corpus study that reported finding Miller's (1956) WM capacity of seven reflected in the word lengths of intonation units. Chafe and Danielewicz (1987), using English corpora, examined one hundred intonation units recorded from natural conversations and one hundred from academic lectures. They found a mean length of 6.2 words in conversation, and 7.3 words in lectures. This mean word length of intonation units, they concluded, was due to the capacity of WM proposed by Miller (1956). Chafe's (1994) follow up work reports only four words per intonation unit on average; yet, even here, he explicitly maintains Miller's (1956) number seven as correct, likely because it had such widespread acceptance at the time. What may have been overlooked in Chafe and Danielewicz (1987) was that other pauses in the speech flow such as hesitations, false starts and repetitions all function to buffer WM

1 The number of canonical clauses and total clauses are different due to spoken language patterns such as: (*T40_8205) Nick, "Still ain't got one"*.

capacity (Ellis and Sinclair 1996). Let us therefore consider the mean length of intonation units again, taking such features into account.

6.1 Corpus consultation: Intonation units

Two corpora were used because unlike the large and balanced BNC, corpora coded for intonation tend to be smaller and unbalanced. One was the Santa Barbara Corpus of Spoken American English (SBCSAE) (Du Bois et al. 2005), which consists of casual spontaneous conversations between two or more speakers. The other was the Talkbank Switchboard Corpus (Graff and Bird 2000), an annotated corpus of 36 telephone conversations between two speakers based on a variety of discussion prompts. These corpora both follow the same discourse analysis transcription system (Du Bois et al. 1993). To estimate the average length of continuous intonation without any pauses or intonation resets in each corpus the following procedures were used. All intonation unit transcription codes (?), (.) and (,), and all other transcribed pauses (..), (...), and (–) were converted to a single code in Notepad ++, namely (.). All transcriber notes and tags were stripped, such as (Hx), <XX>, (cough) etc., and the total word count was then averaged over the single code frequency (.) to produce an estimate of the mean length of continuous intonation. Results are reported in Table 7.

Table 7: Mean length of continuous intonation.

	SBCSAE	Switchboard
Words: mean	3.85	4.03
Cont. intonation: sum	62710	15989
Word count: sum	241525	64569

The mean length of continuous discourse without pauses is around four words in both corpora. The results are consistent with Chafe (1994) and Cowan's (2001) revised WM capacity. It may be that Chafe and Danielewicz (1987) were able to find an intonation unit length consistent with Miller (1956), because by not considering other pauses in intonation, they were looking at intonation units in much the same way as Miller's (1956) experimental conditions measured recall span, which as Cowan (2001) notes, confounded WM capacity by also allowing rehearsal/planning processes to achieve the greater number.

6.2 Spoken Inuktitut

Experimental research has long used the (English) word as a proxy unit in probing WM (Miller 1956; Gilchrist et al. 2008). However, there is no necessary relationship between the word and a WM unit. For example, during language acquisition, syllables and phonemes are a unit of analysis (Kuhl 2004), and even for adult speakers of English in experimental conditions, recall declines with polysyllabic stimuli (Baddeley et al. 1975). Let us therefore consider polysynthetic languages, in which the word is a more complex unit. One suspects that in such languages the patterns would not be identical to the English data above, since the salient information unit to which WM attends might be below the word, perhaps the stress marked root morpheme (Kelly et al. 2014: 60). Inuktitut is an agglutinative language of northern Canada and a member of the Eskimo–Aleut family. An Inuktitut utterance and its English translation are given in (6).

(6) Inuktitut (adapted from Allen and Schröder 2003: 308)
Maunaasijunga
ma una aq si junga
here VIA go PROG PTCP.1sg
I am going through here.

The single word in (6) requires five words in English. This information density is not always true of the language, but if it is correct that WM capacity is not necessarily reflected in word span constraints, then one would expect shorter multi-word sequences in Inuktitut. It would be problematic to find in such languages that intonation units, n-grams and so forth looked the same as in the English or Mandarin data. One would have to argue, in the event of such as finding, that WM always operates on the word, which seems implausible. Such a finding would probably indicate that the above patterns of four above are artifacts of statistics and corpus linguistics, rather than reflecting anything cognitively interesting about WM operations and language.

A spoken corpus of Inuktitut transcribed for intonation was not obtainable, so the Nunavut Hansard Inuktitut-English Parallel Corpus (Martin et al. 2003) was used. This is a corpus of spoken parliamentary proceedings, containing approximately 2.5 million words of Inuktitut. To estimate the length of utterances without pauses, the following procedures were used. All transcription punctuation (,), (.), (?) were converted to a single (.) code; when the transcription lacked utterance final punctuation but this was signaled by a line break, (.) was also added. Using NOTEPAD ++, XML tags were added to all Inuktitut and English utterances, so that the parallel data could be computed separately.

Beside intonation unit length, n-grams were also investigated. Parliamentary language is inherently formulaic, however, typically with every speaker's name and their department mentioned before their utterance. To try to gain a more accurate measure of multi-word sequences in the language, rather than context dependent formulas, all n-grams containing the names of ministers, departments or committees were removed. N-grams were computed in WORDSMITH, with a lower threshold of 40 occurrences p/m as with the BNC. Results are reported in Table 8.

Table 8: Nunavut Hansard Inuktitut-English Parallel Corpus.

Inuk. word count		Inuk. mean IU length		Eng. word count	Eng. mean IU length
2441568		3.35		5480283	5.50
Inuk. n-grams	**2 words**	**3 words**	**4 words**	**5 words**	**6 words**
Types	233	19	0	0	0
Tokens	63338	2639	0	0	0

The data indicates that in the parallel corpus, Inuktitut has a mean IU length of approximately 3.35 words. Since the data is not a narrow transcription that contains all pauses, such as in Section 6.1, one expects that the mean length might in fact be shorter. N-grams are far less frequent in Inuktitut than in English, and there were zero n-grams in Inuktitut of four words or more (though in the parallel English utterances there were 581 types, 153,213 tokens). Even with names and departments removed, many of the two and three word n-grams were still highly contextual. The most frequent two word sequence was *qujannamiik uqaqtii* ('thank you Mr. Speaker'), and inspection of the 19 three word n-grams indicated almost all were parliamentary formula, such as *katimajiralaat katimannanga kajusijunnaqsivuq* ('the committee will come to order'). The findings suggest that the kind of multi-word formulaic language that we have seen in the more analytic English and Mandarin aligning with WM capacity using the word as a proxy item is not in evidence in this language, as expected. Moreover, in a recent review Kelly et al. (2014: 61) notes that although chunk extraction is often cited mechanism in FLA (Pinker 1989; Slobin 1985), it is "not clear what the role of 'chunking' (storing unanalyzed exemplars or chunks) may play in the acquisition of polysynthetic languages … If children do this, we would expect to find learners of languages with complex distributed morphology to be producing ungrammatical unanalyzed chunks on the basis of the speech context – a finding that does not appear strongly in the literature". They conclude that "it remains to be seen what the role of chunking is in other synthetic and polysynthetic languages”. This demonstrates the long way cognitive science has to go before a complete understanding WM, chunking and language.

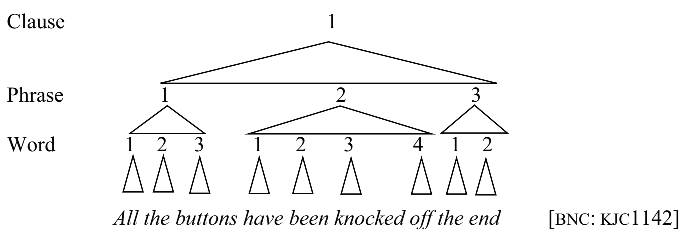
7 Discussion

What has been argued is that a range of structures and patterns in language align with Cowan’s (2001) revised working capacity of four, plus or minus one, something that was not able to be demonstrated for Miller’s (1956) earlier capacity estimate of seven, plus or minus two. The fact that usage-based linguists never found Miller’s (1956) number to have much magic explanatory power is, in retrospect, unusual since WM has been considered a central cognitive domain in psychology for language processing, development and even evolution (Baddeley 2003; Cowan et al. 2005; Hurford 2007). There may be a useful generalization for cognitive linguistics in the above corpus data, followed up of course by further study and experimental evidence: i. e., grammatical patterns are shaped by language use, and language use is constrained by general cognitive domains, so WM limitations come to be reflected in linguistic structure. Consequently, Cowan’s (2001) *magic number four*, plus or minus one, is a (soft) constraint on idioms, phrasal verbs, n-grams, noun phrase length, and even some abstract grammatical constructions.

This paper has looked at abstract grammar that permits fully creative strings, such as phrasal categories and clauses, as well more formulaic strings, such as idioms, phrasal verbs, and n-grams. It has been suggested that both of these reflect WM chunking processes. One question that arises is if, and how, both of these rather different phenomena can be considered chunks. Cowan et al. (2005: 97) argues that not everything is as yet understood about chunking: “at a microscopic level, we are still not sure what is going on and it is only at the intermediate level of analysis that we can draw conclusions about WM capacity in chunks”. The following is therefore a tentative proposal of how chunking applies to both creative language use governed by grammar, and more formulaic language, which may be stored in the lexicon as a more holistic unit (Conklin and Schmitt 2012; Sosa and MacFarlane 2002). Essentially the argument is that they are different types of chunks, but chunks nonetheless, and constrained by the same working memory process. The use of a grammatical construction, which can be fully creative, is a type of associative binding (Gilchrist et al. 2008), in which the chunking operation applies to information that may not previously have been connected (Bergen 2012); formulaic language that is not acquired holistically (in which we might include lexicalized units which underwent associative binding diachronically) are chunks created after repeated experience that have been transferred into LTM (Wray 2002).

Let us first discuss creative language use and grammar. It seems reasonable to suggest that the hierarchical processes of working memory, via chunking, is intimately related to the hierarchical structure of grammar. Pothos (2007: 236) suggests that the emergence of a finite state grammar in artificial grammar learning can be achieved via a generalization across statistical patterns in the input in which "co-occurring chunks or elementary units are organized into larger chunks [and] frequency of occurrence determines the salience of each chunk". An emergent grammar would then consist of constructions that are constrained by WM span, such as in (7).

(7) Hierarchical associative binding by grammar
Clause 1



One might propose the following analysis for (7). WM attends to the sequence of three lower order items *all + the + buttons*, which then constitute an NP in a higher order string. Thus, phrase category grammar binds multiple words into a single unit, thereby refreshing part the WM store (Bergen 2012). This building up of hierarchical structure operates across the VP and NP also creating a sequence of three chunks of lower order items. While some readers may prefer *knocked off* to have a single node or perhaps see the final constituent as a PP, the constituent analysis is not crucial to the general point that rather than WM attending to nine words, grammatical organization binds information into three higher order chunks. In turn, this is organized into an even higher order unit, namely the clause, which typically constitutes a single proposition (of old and new information) in ongoing discourse (Chafe 1994; Givón 2002). This associative binding by hierarchical syntax, from words, to phrasal categories to clauses, allows WM to buffer its limited capacity by organizing lower order items into higher order bindings (Gilchrist et al. 2008). As noted at the outset, not every word in a string has the same cognitive load in WM, but even so psycholinguistic evidence (e. g., from garden path sentences) has shown sentence processing consists of the incremental building up of hierarchal structure so varied informational load of words in a string does not nullify the associative binding process (Clahsen and Felser 2006).

Associative binding likely operates below the word and above the clause also. It has been argued that in FLA the chunking of phoneme sequences leads to syllables, and the chunking of syllables to words (Kuhl 2004; MacWhinney 2005). Incidentally, the current researcher computed a wordlist for the offline version of Corpus of Contemporary American English (142,158 word types) and when processed by a syllable counter (http://countwordsworth.com), the mean length of English words was 2.60 syllables, approximating the WM span of Cowan (2001). At the clause level, the proposition may function as a chunk in discourse, something Chafe (1994) proposed on the basis of corpus data (1 clause approximates 1 intonation unit with 1 new piece of information), and experimentally demonstrated by Gilchrist et al. (2008), who presented sequences of clauses controlled to around four words to (young) adults, and measured retention of these clauses in WM by testing whether participants could recall lexemes in the clause stimuli. They report that participants' recall span was on average 3.27 (*sd*.65) clauses.

A different account may apply to more fixed multi-word sequences, such as one finds in formulaic language. These may already be some form of chunk in the lexicon with processing advantages as a unit, before grammatical operations apply (Bannard and Lieven 2012; Conklin and Schmitt 2012). In language acquisition, Wray (2002, 2013) has proposed a theory known as Needs Only Analysis, which argues a child initially extracts formulaic sequences from the input as single items. These are stored as unanalyzed wholes, and are used holistically or as pivot schemas to achieve communicative routines, such as *I wanna X, more X* (MacWhinney 2005). It is only in later language development that the child breaks down such formulaic language into the grammatical structures and words from which they are made. Highly frequent sequences, Wray (2002, 2013) suggests, may come to be re-chunked in a u-shaped learning process, i. e., via associative binding, but others remain unanalyzed chunk extractions. That is, if there is no need to analyze multi-word sequences as component parts, then this is never done, and language is a mixed system of formulas and 'words and rules'. In relation to the argument in this paper, it may be that fully creative strings use grammar as a chunking mechanism, but formulaic language that has already been chunked by WM via repeated experience (or acquired holistically) need not rely on associative binding for processing advantages. One case would be lexicalized strings such as those detailed in the work of Bybee (2010) which have gone through the process of associative binding diachronically, but once lexicalized constitute a holistic lexical item FLA (i. e., a case of chunk extraction). The general point is that WM may attend to hierarchical syntactic structure on a needs only basis (Wray 2002).

One might ask why formulaic sequences would align in English or Mandarin when chunking four words into one unit would be a type of resetting that might allow continual addition of words to that unit: chunks certainly do build upon other chunks (MacWhinney 2005). Beside WM not being the only explanatory mechanism involved in the linguistic phenomenon discussed in this paper, one possibly that restricts the length of chunks and is related to WM is that the cognitive domain has not only capacity but also resource constraints. As Cowan et al. (2005) notes, WM cognition is constrained by time, which limits processing to around 2 seconds. It is also constrained by energy, which limits the complexity of units that can be attended to given physiological processes such as neuron activation. Perhaps chunks up to 12 words (i. e., associatively bound 4 words x 4 chunks) are simply too complex and take too long to produce. As Cowan et al. (2005) mention, however, the interactions of WM capacity and other processes can make it difficult to settle on an explanatory variable. Conversely, one can also ask why if up to four words can be chunked, there are sequences with fewer – might it not be maximally efficient to chunk all information strings close to capacity and store these in long term memory? Perhaps, but the WM research does not suggest equal ease of processing for one versus four items. And, other factors such as frequency, semantic coherence and discourse functionality are also highly relevant. For example, even with largely statistical phenomenon such as n-grams, Biber et al. (2004) demonstrated that they tend to have relatively clear discourse functions and taxonomies. If a two word recurrent sequence is a coherent unit and high frequency, it may be chunked into LTM.

Relatedly, an anonymous reviewer suggested, the corpus data on phrasal verbs, idioms, n-grams, and noun phrase length is not inconsistent with a smaller number than Cowan's (2001); perhaps a magic number three, plus or minus one. While Cowan's (2001) estimate has by now much supporting experimental evidence, smaller capacities have indeed been proposed. For example, Broadbent (1975) proposed three as the capacity of working memory, and Gobet and Clarkson (2004) suggest Cowan's (2001) number may be an overestimate by one or perhaps two. Oberauer (2005) has even argued that WM only has a capacity of one, with a periphery of 3–5. Cowan et al. (2005: 58) responds to such alternative estimates by arguing that the same evidence can be explained by the 3–5 model, but notes substantial variation "with a range possibly spreading from 2–6 chunks in normal adults". Further, some cognitive scientists argue for non-capacity memory models such as Van Dyke and Johns (2012) who suggest that what may look like capacity constraints could results from resource constraints, interference and decay. It may turn out that another revision to WM occurs as cognitive science progresses.

8 Limitations and future research

This study needs to be followed with a series of studies into the interaction between WM and linguistic structure. Experimental evidence is needed, as behavioral observation data from corpora is not enough on its own. The relationship between formulaic chunks and associative binding via grammar needs to be fully been worked out, along with other variables such as the role of metrical structure, morphological complexity etc. Technical limitations of this paper include a margin of error in counts. N-gram computation detects overlapping sequences, such as *at the end of the day, at the end of the, the end of*, and so forth; the script for NP modification was imperfect. It is not known how representative the web corpora of idioms and phrasal verbs are; dictionaries vary with respect to the examples they include. It should also be noted that the large scale matching procedure in the BNC was insensitive to which phrasal verbs and which idioms were literal uses, i. e., the counts do not disambiguate literal and idiomatic uses of *kick the bucket*. A match-list built from dictionary headwords did not find phrasal verbs or idioms that are variable: *he put on a show* was counted but not *he put a show on*. Finally, in the analysis of intonation, repetition was not counted due to the difficulty in obtaining counts, but hesitations such *I I I don't know* are certainly buffers to WM (Ellis 2001).

9 Conclusion

Usage-based linguistics never found Miller's (1956) *magic number seven, plus or minus two,* to be a recurrent feature in the grammatical patterns of language. This is perhaps not surprising, since the WM capacity proposed by Miller (1956) appears to be an overestimate after many years of such wide acceptance, and it was during these years that usage-based linguistics developed as a branch of cognitive science. Cowan's (2001) seminal revision of WM capacity to a *magic number four, plus or minus one*, has been a major development in cognitive science in the past 15 years, and the current study has demonstrated an alignment in corpora with this revised WM capacity. The paper has opened the discussion space for what may be an important cognitive constraint for usage-based linguistics to attend to. The revised WM capacity has been given little attention thus far in linguistics, unlike the other branches of cognitive science. This is a significant area for future research which can promote a more complete integration of cognitive linguistics into cognitive science.

Acknowledgements: The author would like to thank Dr. James Lambert for his assistance with a Perl script for NP extraction. Further, I am grateful to Dr. Mike Scott for some technical advice, and for the immensely valuable input of the anonymous reviewers and editorial team.

References

Aboitiz, Francisco, Ricardo R. García, Conrado Bosman & Enzo Brunetti. 2006. Cortical memory mechanisms and language origins. *Brain and Language* 98. 40–56.

Allen, Shanley & Heike Schröder. 2003. Preferred argument structure in early Inuktitut spontaneous speech data. In John Du Bois, Lorriane Kumpf & William Ashby (eds.), *Preferred argument structure: Grammar as architecture for function*, 301–338. Amsterdam: John Benjamins.

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

Baddeley, Alan. 2003. Working memory and language: An overview. *Journal of Communication Disorders* 36. 189–208.

Baddeley, Alan. 2007. *Working memory, thought, and action*. Oxford: Oxford University Press.

Baddeley, Alan, Neil Thomson & Mary Buchanan. 1975. Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior* 14(6). 575–589.

Bannard, Colin & Elena Lieven. 2012. Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics* 32. 3–16.

Bannard, Colin & Danielle Matthews. 2008. Stored word sequences in language learning the effect of familiarity on children's repetition of four-word combinations. *Psychological Science* 19. 241–248.

Bergen, Benjamin. 2012. *Louder than words: The new science of how the mind makes meaning*. Philadelphia: Basic Books.

Biber, Douglas, Susan Conrad & Vivian Cortes. 2004. If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3). 371–405.

BNC – *The British National Corpus*, version 3 (BNC XML edn.). 2007. Distributed by Oxford University Computing Services on behalf of the BNC Consortium. http://www.natcorp.ox.ac.uk/p

Bor, Daniel. 2016. Advances in the scientific investigation of consciousness. In Martin Monti & Walter G. Sannita (eds.), *Brain function and responsiveness in disorders of consciousness*, 13–24. Cham: Springer.

Broadbent, Donald. 1975. The magic number seven after fifteen years. In Alan Kennedy & Alan Wilkes (eds.), *Studies in long-term memory*, 2–18. New York: Wiley.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4). 711–733.

Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
Chafe, Wallace. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press.
Chafe, Wallace & Jane Danielewicz. 1987. Properties of spoken and written language. In Rosalind Horowitz & S. Jay Samuels (eds.), *Comprehending oral and written language*, 83–113. New York: Academic Press.
Clahsen, Harald & Claudia Felser. 2006. Grammatical processing in language learners. *Applied Psycholinguistics* 27(1). 3–42.
Collins, Peter & Carmella Hollo. 2009. *English grammar: An introduction*. Chippenham: Palgrave McMillian.
Colston, Herbert. 2008. A new look at common ground: Memory, egocentrism, and joint meaning. In Istvan Kecskes & Jacob Mey (eds.), *Intention, common ground and the egocentric speaker-hearer*, 151–187. Berlin: Walter de Gruyter.
Conklin, Kathy & Norbert Schmitt. 2012. The processing of formulaic language. *Annual Review of Applied Linguistics* 32. 45–61.
Cowan, Nelson. 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24(1). 87–185.
Cowan, Nelson, Emily Elliott, Scott Saults, Candice Morey, Sam Mattox, Anna Hismjatullina & Andrew Conway. 2005. On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology* 51(1). 42–100.
Cowan, Nelson, Candice Morey, Zhijian Chen, Amanda Gilchrist & Scott Saults. 2008. Theory and measurement of working memory capacity limits. *Psychology of Learning and Motivation* 49. 49–104.
Denison, David. 2004. *English historical syntax: Verbal constructions*. London: Longman.
Downing, Angela. 2015. *English grammar: A university course*. London: Routledge.
Du Bois, John, Wallace Chafe, Charles Meyer, Sandra Thompson & Nii Martey. 2005. *Santa Barbara corpus of spoken American English, Parts 1–4*. Philadelphia: Linguistic Data Consortium.
Du Bois, John, Stephan Schuetze-Coburn, Susanna Cumming & Danae Paolino. 1993. Outline of discourse transcription. In Jane Edwards & Martin Lampert (eds.), *Talking data: Transcription and coding in discourse research*, 45–89. Hillsdale, NJ: Erlbaum.
Ellis, Nick. 2001. Memory for language. In Catherine Doughty & Paul Robinson (eds.). *Cognition and second language instruction*, 33–68. Cambridge: Cambridge University Press.
Ellis, Nick & Susan Sinclair. 1996. Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *The Quarterly Journal of Experimental Psychology* 49(1). 234–250.
Ericcson, K. Anders, William Chase & Steven Faloon. 1980. Acquisition of a memory skill. *Science* 208. 1181–1182.
Evans, Nicholas & Stephen C. Levinson. 2009. The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences* 32(5). 429–448.
Evans, Vyvyan. 2014. *The language myth: Why language is not an instinct*. Cambridge: Cambridge University Press.
Gathercole, Susan & Tracy Alloway. 2008. *Working memory and learning: A practical guide for teachers*. Sage: London.
Gathercole, Susan, Leanne Brown & Susan Pickering. 2003. Working memory assessments at school entry as longitudinal predictors of national curriculum attainment levels. *Educational and Child Psychology* 20(3). 109–122.

Gazzaniga, Michael. 2015. *Tales from both sides of the brain: A life in neuroscience*. New York: HarperCollins.
Gernsbacher, Morton Ann & Talmy Givón (eds.). 1995. *Coherence in spontaneous text*, Vol. 31. Amsterdam: John Benjamins.
Gilchrist, Amanda, Nelson Cowan & Moshe Naveh-Benjamin. 2008. Working memory capacity for spoken sentences decreases with adult ageing: Recall of fewer but not smaller chunks in older adults. *Memory* 16(7). 773–787.
Givón, Talmy. 1984. Prolegomena to discourse-pragmatics. *Journal of Pragmatics* 8(4). 489–516.
Givón, Talmy. 2002. *Bio-linguistics: The Santa Barbara lectures*. Amsterdam: John Benjamins.
Givón, Talmy. 2012. The adaptive approach to grammar. In Bernd Heine & Heiko Narrog (eds.), *The Oxford handbook of linguistic analysis*, 27–51. Oxford: Oxford University Press.
Gobet, Fernand & Gary Clarkson. 2004. Chunks in expert memory: Evidence for the magical number four … or is it two? *Memory* 12(6). 732–747.
Goldberg, Adele. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
Graff, David & Steven Bird. 2000. Many uses, many annotations for large speech corpora: Switchboard and TDT as case studies. In Maria Gavrilidou, George Carayannis, Stelios Piperidis, Stella Markantonatou & Gregory Steinhauer (eds.), *Proceedings of the Second International Conference on Language Resources and Evaluation*, 427–433. Athens: LREC.
Haspelmath, Martin. 2003. Explaining the ditransitive person-role constraint: A usage-based approach. *Constructions* 2. 1–71.
Haspelmath, Martin. 2008. Frequency vs. iconicity in explaining grammatical asymmetries. *Cognitive Linguistics* 19(1). 1–33
Henry, Lucy. 2012. *The development of working memory in children*. London: SAGE.
Hills, Peter & Michael Pake. 2016. *Cognitive psychology for dummies*. Chichester: Wiley.
Hilpert, Martin. 2014. *Construction grammar and its application to English*. Edinburgh: Edinburgh University Press.
Huddleston, Rodeny & Geoffrey Pullum 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
Hurford, James. 2007. *The origins of meaning: Language in the light of evolution*, Vol. 1. Oxford: Oxford University Press.
Ibbotson, Paul & Michael Tomasello. 2009. Prototype constructions in early language acquisition. *Language and Cognition* 1(1). 59–85.
Jackendoff, Ray. 2003. Précis of foundations of language: Brain, meaning, grammar, evolution. *Behavioral and Brain Sciences* 26(6). 651–665.
Kelly, Barbara, Gillian Wigglesworth, Rachel Nordlinger & Joe Blythe. 2014. The acquisition of polysynthetic languages. *Language and Linguistics Compass* 8(2). 51–64.
Kuhl, Patricia. 2004. Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience* 5. 831–843.
Lakoff, George. 2009. The neural theory of metaphor. In Raymond W. Gibbs Jr. (ed.), *The Cambridge handbook of metaphor and thought*, 17–38. Cambridge: Cambridge University Press.
Langacker, Ronald. 2009. Cognitive (construction) grammar. *Cognitive Linguistics* 20(1). 167–176.
Levelt, Willem, Ardi Roelofs & Antje Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22(1). 1–38.

Liu, Dilin. 2003. The most frequently used spoken American English idioms: A corpus analysis and its implications. *Tesol Quarterly* 37(4). 671–700.
MacWhinney, Brian. 2005. A unified model of language acquisition. In Judith Kroll & Annette De Groot (eds.), *Handbook of bilingualism: Psycholinguistic approaches*, 39–67. New York: Oxford University Press.
Mahlberg, Michaela, Kathy Conklin & Marie-Josée Bisson. 2014. Reading Dickens's characters: Employing psycholinguistic methods to investigate the cognitive reality of patterns in texts. *Language and Literature* 23(4). 369–388.
Martin, Joel, Howard Johnson, Benoit Farley & Anna Maclachlan. 2003. Aligning and using an English-Inuktitut parallel corpus. In *Proceedings of the HLT-NAACL 2003 Workshop on Building and using parallel texts: Data driven machine translation and beyond*, 115–118. Edmonton: Association for Computational Linguistics.
Martinez, Ron & Norbert Schmitt. 2012. A phrasal expressions List. *Applied Linguistics* 33(3). 299–320.
Miller, George. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63(2). 81–97.
Oberauer, Klaus. 2005. Control of the contents of working memory – a comparison of two paradigms and two age groups. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 31(4). 714–728.
Pinker, Steven. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.
Pothos, Emmanuel. 2007. Theories of artificial grammar learning. *Psychological Bulletin* 133(2). 227–244.
Pothos, Emmanuel & Patrick Juola. 2001. Linguistic structure and short term memory. *Behavioral and Brain Sciences* 24(1). 138–139.
Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartik, 1985. *A comprehensive grammar of the English language*. London: Longman.
Robinson-Riegler, Bridget & Gregory Robinson-Riegler. 2008. *Cognitive psychology: Applying the science of the mind*. London: Pearson.
Sampson, Geoffrey, Abdul Rahman & Alanna Morris. 2000. *The Christine corpus (Release 2)*. Sussex: University of Sussex.
Schmitt, Norbert, Sarah Grandage & Svenja Adolphs. 2004. Are corpus-derived recurrent clusters psycholinguistically valid? In Norbert Schmitt (ed.), *Formulaic sequences: Acquisition, processing and use*, 127–151. Amsterdam: John Benjamins.
Scott, Mike. 2015. WordSmith Tools (Version 6) [software]. Liverpool.
Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.
Siyanova-Chanturia, Anna. 2015. On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory* 11. 285–301.
Slobin, Dan I. (ed). 1985. *The crosslinguistic study of language acquisition: Theoretical issue*, Vol. 2. New Jersey: Erlbaum.
Sosa, Anna Vogel & James MacFarlane. 2002. Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language* 83(2). 227–236.
Van Dijk, Teun & Walter Kintsch. 1983. *Strategies of discourse comprehension*. New York: Academic Press.
Van Dyke, Julie A. & Clinton L. Johns. 2012. Memory interference as a determinant of language comprehension. *Language and Linguistics Compass* 6(4). 193–211.

Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.

Wray, Alison. 2013. Formulaic language. *Language Teaching* 46(3). 316–334.

Wu, Chu-hsia. 1995. On the cultural traits of Chinese idioms. *Intercultural Communication Studies* 5. 61–82.

Xiao, Richard. 2010. Idioms, word clusters and reformulation markers in translational Chinese. In Richard Xiao (ed.), *Proceedings of the international symposium on using corpora in contrastive and translation studies* 2010 *conference* (UCCTS2010). Lancaster: Lancaster University.