Evidence for a facilitatory effect of multi-word units on child word learning

Robert Grimm, **Giovanni Cassani**, **Walter Daelemans** and **Steven Gillis**
Computational Linguistics and Psycholinguistics (CLiPS) Research Center
Department of Linguistics, University of Antwerp, 13 Prinsstraat
B-2000 Antwerpen, Belgium
{name.surname}@uantwerpen.be

Abstract

Previous studies have suggested that children possess cognitive representations of multi-word units (MWUs) and that MWUs can facilitate the acquisition of smaller units contained within them. We propose that the formation of MWU representations precedes and facilitates the formation of single-word representations in children. Using different computational methods, we extract MWUs from two large corpora of English child-directed speech. In subsequent regression analyses, we use age of first production of individual words as the dependent and the number of MWUs within which each word appears as an independent variable. We find that early-learned words appear within many MWUs – an effect which is neither reducible to frequency or other common co-variates, nor to the number of context words contained in the MWUs. Our findings support accounts wherein children acquire linguistic patterns of varying sizes, moving gradually from the discovery of MWUs to the acquisition of small-grained linguistic representations.[1]

Keywords: multi-word units; age of first production; word learning; language acquisition; computational modeling

Introduction

Frequently co-occurring word combinations have been investigated in studies examining both child (Bannard & Matthews, 2008; Arnon & Clark, 2011; McCauley & Christiansen, 2014) and adult processing (Arnon & Snider, 2010), with mounting evidence that children and adults represent such sequences separately from their constituent words. Indeed, given that many English word sequences have idiosyncratic meanings which cannot be derived from the meaning of their constituent words (e.g. *pay attention to*, *leave of absence*, *you're welcome*), it is reasonable to expect language users to store such semantically opaque sequences in memory. Findings from the literature, however, extend beyond this: in addition to non-compositional constructions, people are likely to also lexicalize frequent but semantically transparent formulaic sequences (Wray, 2008). Here, we use the term *multi-word unit* (MWU) to refer to any sequence of words – compositional or not – which is likely to be lexicalized, and we investigate the role of MWUs in the acquisition of individual words.

More concretely, we expect a facilitatory interaction between the acquisition of MWUs and the acquisition of their constituent words. Provisional evidence for a beneficial impact of MWUs on the acquisition of smaller Linguistic units was collected by Arnon and Clark (2011), who showed that children make fewer inflectional errors on known words if the words are contained within frequent MWUs. Usage-based approaches to language acquisition, meanwhile, suggest that children acquire a repertoire of both lexically specific and more abstract multi-word constructions (Tomasello, 2009; Behrens, 2009). Based on this, we propose that children sometimes possess MWU representations before they form representations of the words contained within them, and that these MWU representations then facilitate the acquisition of single-word representations. We dub this the *MWU acquisition hypothesis*.

With the availability of a growing number of corpora of child-caregiver interactions on the one hand (MacWhinney, 2000) and the development of methods for the extraction of MWUs from corpora on the other hand (McCauley & Christiansen, 2014; Brooke, Tsang, Hirst, & Shein, 2014), we are in a position to investigate the kinds of MWUs children are likely to acquire. Concretely, we extract MWUs from two large corpora of transcribed child-directed speech, using (a) a computational model employed by McCauley and Christiansen (2014) to account for findings from the language acquisition literature and (b) an algorithm, developed by Brooke et al. (2014), intended to build a comprehensive lexicon of psychologically plausible MWUs. We view extracted MWUs as an approximation of the types of MWUs children might discover and use the number of MWUs within which a given word is contained as an independent variable.

Throughout, we use the age at which children first produce words (age of first production / AoFP) as an index of word learning: if a word is first produced relatively early in development, we assume that this is in part because it is easy to learn when and how to use it. Given the *MWU acquisition hypothesis*, we expect a facilitatory effect of the number of MWUs in which a word appears on its AoFP. This effect, moreover, should be uniquely attributable to MWUs – and not to individual word frequency, semantic co-variates, or the number of context words contained in MWUs. Number of co-occurring context words has previously been shown to predict age of acquisition for words (Hills, Maouene, Riordan, & Smith, 2010); but if our proposal is correct, such an effect should disappear once MWUs are taken into consideration.

Related Work

Language Acquisition

MWUs have emerged as an important theoretical concept in usage-based approaches to Language Acquisition (Tomasello, 2009). Within this broad theoretical framework,

[1] The code for running our experiments is available online: https://github.com/RobGrimm/CogSci2017-MultiWordUnits

learners' linguistic representations are conceived of as continually complexifying entities, with the developed cognitive system containing both lexically specific and more abstract patterns. At early stages in development, most representations are lexically specific, and child language is "(partially) formulaic and item-based" (Behrens, 2009, p. 393). That is, child language development is thought to involve representations which are lexically specific and span multiple words.

Experimental evidence for the existence of children's MWU representations comes from Bannard and Matthews (2008), who presented 2 and 3 year-olds with frequent MWUs like *a drink of tea* and matched infrequent MWUs like *a drink of milk* that differed in the last word. 2 and 3 year-olds were faster to repeat frequent MWUs, and 3 year-olds were also faster to repeat the first three words if they formed a frequent MWU with the fourth word. Since the final word and the final bigram (e.g. *of tea* and *of milk*) were matched for frequency, the processing advantage for frequent MWUs can only be attributed to the frequency of the entire MWU, rather than to the frequencies of its component words, suggesting that children have access to cognitive representations of MWUs. Bannard and Matthews (2008) argue, furthermore, that their subjects were likely familiar with the words comprising the MWUs, which implies the existence of (partially) independent MWU and single-word representations.[2]

In addition, Arnon and Clark (2011) found that MWUs interact with the acquisition of morphemes: 4;6 year-olds produced more correct irregular plurals after familiar lexically specific frames than after general questions. Subjects were presented with depictions of several objects. The object name was elicited either with a labeling question or with a lexically specific frame. For example, on one particular trial the objects were sheep, the lexically specific frame was *Count some* –, and the labeling question was *What are all these called?* 4;6 year-olds were more likely to complete the lexically specific frame with *sheep* and would provide relatively more incorrect plural forms – like the over-regularized *sheeps* – in response to the labeling question. This suggests that MWUs like *count some sheep* affect the way in which some of the smaller units contained within them are learned.

Computational Modeling

The above-cited results by Arnon and Clark (2011) and Bannard and Matthews (2008) have been modeled by McCauley and Christiansen (2014). In a comprehension phase, their model segments a corpus of child-directed speech into MWUs. In a production phase, it generates child-produced utterances based on stored MWUs. Given a corpus, MWUs are extracted by comparing the conditional probability of the current word given the preceding word to a running average of all such probabilities, for all words so far encountered one position to the left of the current word. If this backward transitional probability (BTP) is larger than the running average, the current and preceding word are part of an MWU. The process continues until the BTP falls below the average, at which point the current MWU is stored in memory.

Extracted MWUs can then be used to re-construct child-produced utterances. McCauley and Christiansen (2011) compared model-derived to child-produced utterances across 13 corpora from the CHILDES database (MacWhinney, 2000). On average, about 60 % of utterances were successfully re-produced – illustrating that a purely MWU-based system can account for a majority of child-produced utterances. Importantly, MWUs discovered by the model can also be used to model results from Bannard and Matthews (2008) and Arnon and Clark (2011). In both cases, stimuli were sequences of words – constructions like *a drink of tea* in the former and *count some sheep* in the latter study. McCauley and Christiansen (2014) assigned a *chunkedness* score to each stimulus by calculating the product of BTPs between the MWUs used by the model to re-produce each stimulus. In each study, differences in scores reflected differences in subjects' performance: stimuli with lower reaction times in Bannard and Matthews (2008)'s study were assigned a larger chunkedness score, as were stimuli which elicited a larger proportion of correctly inflected nouns in Arnon and Clark (2011)'s study.

Natural Language Processing

McCauley and Christiansen's (2011, 2014) model can be situated in a tradition that measures association strength between pairs of words; words are then grouped together if their association strength exceeds a particular threshold. McCauley and Christiansen (2014, 2011) use BTP as the measure of association. Other options include pointwise mutual information or log likelihood (cf. Pecina, 2010, for an overview). All association-based methods require an arbitrary threshold for inclusion of words in MWUs. In addition, there is no consensus on which association measure is best. An alternative approach is to identify frequent n-grams – called *lexical bundles* –, but this requires very high frequency thresholds (Biber, Conrad, & Cortes, 2004). There is, then, no generally accepted way of extracting MWUs from corpora, nor is it common practice to evaluate whether extracted MWUs correspond to psychologically real entities.

Work by Brooke et al. (2014) has recently begun to address these issues. Their method operates at the token level, identifies MWUs of varying sizes, and relies on two parameters: a frequency threshold and a maximum MWU size. Broadly speaking, the algorithm considers all possible segmentations of a given sentence into n-grams that meet a pre-specified frequency threshold. Then, that segmentation is selected which maximizes the predictability of each word within its n-gram. The stated goal of this work is to develop a method for the extraction of an MWU lexicon that would correspond to knowledge of MWUs possessed by native speakers. The system has since been refined by Brooke et al. (2015), who also in-

[2] The same argument can be made for adults, who are faster to recognize and produce frequent four-word MWUs in similar experiments (Arnon & Snider, 2010). Such results also support theories of adult linguistic competence which include MWU-like constituents (O'Donnell, 2015).

troduced first steps towards evaluating MWU lexicons.

Hypothesis

According to the *MWU acquisition hypothesis*, children sometimes acquire MWU representations before they acquire representations of the individual words contained in MWUs, and access to MWU representations then facilitates acquisition of the words contained in them.[3] While this hypothesis is grounded in the literature, it is not clear via which mechanisms MWUs might aid the word learning process. Consequently, our goal is to provide evidence *that* MWUs uniquely facilitate word learning, and not *how* this process unfolds. Below, we nevertheless sketch two possible scenarios.

One possibility is that children initially acquire MWUs as unanalyzed units. This could result from an initial undersegmentation of the input: words, before their meaning is established, need to be identified from a continuous stream of sound. Early in development, children might sometimes segment multi-word chunks before they begin to segment individual words from within those chunks. Thus, some early fossilized MWUs are likely to be (partially) undersegmented chunks. In this scenario, the more initially undersegmented MWUs contain it, the earlier a given word is going to be segmented. We would then expect this early segmentation to translate into early induction of meaning.

A second possibility is that children discover some words before establishing their meaning. They would then go on to discover MWUs containing those words, at which point they have access to fully-fledged MWU representations without having access to the meaning of each individual word. The more MWUs contain a given word, the more words it is going to be linked to – and the more words will prime its retrieval, making it more salient for the learner. On average, a word with many links will be more easily retrieved than a word with few links. Because of this, we would expect fewer necessary exposures to establish the meaning of a word which forms part of relatively many MWUs, compared to words contained in fewer MWUs.

As mentioned, we do not distinguish between these two and other such possibilities. Instead, we aim to broadly corroborate the *MWU acquisition hypothesis* by showing that MWUs uniquely facilitate word learning: if, all else being equal, words contained in many MWUs are learned earlier than other words, this would be indicative of a developmental pattern which begins with the formation of MWU representations and then proceeds to the acquisition of individual words.

Method

Our method is the following: first, we extract MWUs from two corpora of English child-directed speech (CDS) and estimate age of first production (AoFP) for the words produced by the children addressed in the CDS corpora. We then use the number of MWUs within which each target word appears (*#MWUs*) as an independent variable – next to several covariates – in a linear regression analysis, with AoFP as the dependent variable. If the *MWU acquisition hypothesis* is true, we expect a unique facilitatory effect of *#MWUs* on AoFP.

Child-Directed Speech

We use two corpora of CDS, which both consist of the adult-produced utterances from several corpora on the CHILDES database (MacWhinney, 2000). Some corpora are based on cross-sectional studies, while others are longitudinal. In addition, subjects vary in age. Regardless, each corpus consists of standardized transcripts, based on recordings of child-caregiver interactions. In order to maximize the amount of data, we ignore possible fine-grained differences between age cohorts and compile a North-American corpus (NA-CDS) from 45 American English corpora[4] and a British English corpus (BE-CDS) from eight British corpora[5]. Table 1 summarizes statistics.

Table 1: Relevant corpus statistics.

measure	CDS-BE	CDS-NA
nr. tokens	4,681,925	6,389,963
nr. types	24,929	37,128
median length of utt.	4 (IQR: 4)	4 (IQR: 4)
nr. adult speakers	201	774
nr. children addressed	134	441
mean child age (months)	33 (SD: 9)	41 (SD: 23)

Extraction of Multi-Word Units

To extract MWUs from the CDS corpora, we use McCauley and Christiansen's (2014) model as well as Brooke et al.'s (2014) method. McCauley and Christiansen's (2014) model – called *Chunk-Based Learner* (CBL) – processes a given corpus utterance by utterance and word by word. Processing an utterance u is initiated by incrementing the frequency count of the first word $w_1 \in u$ by 1 and creating a new MWU with w_1 as its only member. For each subsequent word w_i at utterance position $1 < i \leq length(u)$, the model keeps track of the number of times w_i has been encountered so far, as well as how often the immediately preceding word w_{i-1} has occurred one position to the left of w. The model then calculates the backward transitional probability (BTP) of w_i and w_{i-1}: $p(w_{i-1}|w_i)$. If this probability is larger than the average BTP across all words which have occurred one position to the left

[3] Note that we do not claim that the acquisition of MWUs *always* precedes the acquisition of single words, but merely that this happens often enough to have a measurable impact on word learning.

[4] Corpora names (see http://childes.talkbank.org/access/ for references): Bates, Bernstein, Bliss, Bloom70, Bloom73, Bohannon, Braunwald, Brent, Brown, Carterette, Clark, Cornell, Demetras1, Demetras2, ErvinTripp, Evans, Feldman, Garvey, Gathercole, Gleason, HSLLD, Hall, Higginson, Kuczaj,MacWhinney, McCune, McMillan, Morisset, Nelson, NewEngland, Peters, Post,Providence, Rollins, Sachs, Snow, Soderstrom, Sprott, Suppes, Tardif, Valian, VanHouten, VanKleeck, Warren, Weist

[5] Belfast, Fletcher, Manchester, Thomas, Tommerdahl, Wells, Forrester, Lara

of w in all utterances so far considered, w_i is added to the current MWU. Else, the current MWU is added to a set M, and a new MWU is created – again with w_i as its only member. In this way, the model discovers MWUs of size 2 or larger, as well as single-word units, collected in M. In our analyses, we use all MWUs which occur at least twice in the input corpus.

As a second model, we use the method from Brooke et al. (2014)[6]. We refer to it as *Prediction Based Segmenter* (PBS), as it splits utterances into n-grams whose component words are maximally predictable. The basic idea is that given an n-gram $w_1 ... w_n$, the conditional probability of any word w_i given the remaining subsequence $w_1 ... w_{i-1}, w_{i+1} ... w_n$ should be maximal. In essence, the algorithm splits utterances into n-grams such that each word's predictability is maximized, capturing the intuition that words within MWUs are more predictive of one another than words outside of MWUs – but see Brooke et al. (2014) for a more in-depth explanation. Specifying a maximum n-gram length of ten – longer than most utterances in the corpus – , we use the PBS to segment utterances into either single-word units or MWUs with a minimum size of two and a maximum size of ten. As with the CBL, we retain all MWUs which occur at least twice.

Running the models on the two CDS corpora results in four different sets of MWUs, whose distributions are summarized in Table 2. The CBL results in a larger number of shorter MWUs, while the PBS identifies MWUs that are a bit longer. There are generally more MWU types than word types (compare Table 1).

Table 2: Relevant statistics about the distribution of MWUs.

corpus	*measure*	CBL	PBS
CDS-BE	MWU tokens	1,073,037	978,804
	MWU types	465,447	387,391
	median length	4 (IQR: 3)	5 (IQR: 4)
CDS-NA	MWU tokens	1,40,8614	1,338,173
	MWU types	628,252	492,863
	median length	4 (IQR: 3)	5 (IQR: 4)

Age of First Production

To induce AoFP, we start from a corpus of child-produced utterances, treating a word as having been learned at the earliest developmental stage at which any child within the corpus can produce it. *Developmental stage* is defined in terms of mean length of utterance (MLU) – the average child utterance length, in tokens, within a transcript. Since transcripts have varying lengths, we estimate MLU for each transcript via statistical bootstrapping, wherein the sampling distribution of the population is approximated by drawing random samples from the data (Davison & Hinkley, 1997). Each bootstrap is based on 1000 random samples with replacement, with the sample size equal to the number of child utterances per transcript. We thus induce MLU rather than AoFP estimates, since MLU is a more robust estimator of development (Parker & Brorson, 2005): children who are close in age may nevertheless be far apart in terms of language development. For simplicity, we still refer to a word's MLU value as its AoFP. To induce a value for any word, we calculate the set of MLUs γ for all transcripts within which the word appears and assign it the smallest value in γ.

We perform this procedure for each word produced by the children addressed in the two CDS corpora – once for the NA data and once for the BE data, meaning that we end up with two AoFP data sets: 441 children are addressed in the CDS-NA corpus and together produce 29,188 different words, each of which is assigned an AoFP value; and 134 children are addressed in the CDS-BE corpus, producing 14,747 different words, again each with its own AoFP value.

Regression Analyses

In the regression models, we use AoFP as the dependent variable. The first key independent variable is the number of different MWUs within which a given target word appears (**#MWUs**). For example, assuming our corpus is CDS-NA and our target words are *girl* and *sit*, we count the unique MWUs which contain these two words. To illustrate this, Table 3 shows the five most frequent MWUs, in CDS-NA, containing the two words. Counting all such MWUs, we end up with 113 (PBS) and 230 MWUs (CBL) for *girl*, and 253 (PBS) and 488 (CBL) MWUs for *sit*. The second key independent variable is the number of unique context words appearing in all MWUs within which a given target word is contained (**#ctxt**). If MWUs aid word learning, we should see a facilitatory effect of *#MWUs* on AoFP, and this effect should not be reducible to *#ctxt*. If a target word appears within a large number of MWUs, it will also tend to co-occur with a large number of context words. We posit, however, that MWUs – not individual words – are the cognitively relevant units; and we predict, therefore, that it is the number of MWUs – not the number of co-occurring context words – which affects learning.

Further, we include the following co-variates: the corpus-frequency of each target word (**freq**), number of syllables (**syl**), phonological neighborhood density (**phon**), and concreteness ratings (**con**). Given a target word, *phon* is defined as the number of homophones, plus the number of words that can be derived from the target by either adding, deleting, or substituting a single phoneme. *phon*, together with *nsyl*, is derived from the CMU pronunciation dictionary[7]. Concreteness ratings for 40,000 lemmas are taken from Brysbaert, Warriner, and Kuperman (2014)[8], who collected them from over four thousand participants via Mechanical Turk. Since ratings were collected for lemmas, whereas we work with word forms, we assigned the lemma rating to all word forms which correspond to the lemma. Regression analyses are based on

[6]available online: http://www.cs.toronto.edu/~jbrooke

[7]http://www.speech.cs.cmu.edu/cgi-bin/cmudict

[8]http://crr.ugent.be/archives/1330

Table 3: The five most frequent MWUs, found in CDS-NA, for the target words *girl* and *sit*. Frequency counts for the MWUs are give in parentheses.

word	CBL	PBS
girl	good girl (410)	good girl (440)
	little girl (110)	little girl (175)
	that's a girl (101)	a girl (98)
	a girl (68)	that's a good girl (59)
	that's a good girl (57)	the little girl (51)
sit	sit down (627)	sit down (846)
	sit up (88)	sit up (141)
	sit here (46)	you sit (117)
	sit over here (46)	you wanna sit (87)
	sit down please (41)	come sit (85)

all words for which *phon*, *syl* and *con* estimates are available: 7,265 words in CDS-BE and 5,724 words in CDS-NA. Table 4 shows three example data points.

To increase the generality of this study's implications, we use AoFP from children who were not addressed in the corpus used to estimate *#MWUs*, *#ctxt*, and frequency. In other words, we use AoFP from the children addressed in the CDS-NA corpus for regression models which include *#MWUs*, *#ctxt* and frequency counts from CDS-BE; and we use AoFP from the children addressed in CDS-BE for regression models which include independent variables from CDS-NA.

Table 4: Example data points from the CDS-BE corpus, with *#MWUs* and *#ctxt* estimated via the PBS. The *phon* and *nsyl* predictors are not shown due to space constraints.

word	freq	con	#ctxt	#MWUs	AoFP
goes	3,183	2.19	430	156	0.51
lunch	1,175	4.31	168	57	1.29
running	853	4.27	86	46	1.16

Results

Table 5 presents results of four linear regression analyses (2 methods for MWU extraction $\times$ 2 CDS corpora). All variables are log-transformed, and *#ctxt* as well as *#MWUs* are increased by 1, in order to avoid problems from zero counts. The baseline models with all co-variates (second column) explain between 38 and 44 percent of the variance in AoFP. *Freq* and *con* have facilitatory effects, while there are no statistically significant effects for *phon* and *nsyl*. Given that increased frequency of exposure is associated with early word learning (Ambridge, Kidd, Rowland, & Theakston, 2015), the effect of *freq* is not surprising, while the effect of *con* implies that words associated with concrete concepts tend to be early-acquired.

Adding *#ctxt* to the baseline models (third column) leads to a significant increase in R^2, with a facilitatory effect of *#ctxt*. Adding *#MWUs* to the baseline models (fourth column) also improves the fit, with a facilitatory effect of *#MWUs*. Interestingly, the effect of *#MWUs* is stronger than the effect of *#ctxt*. Neither effect is reducible to the frequency of target words, their concreteness, their phonological complexity, or the density of their phonological neighborhoods. In models which include the covariates plus *#ctxt* and *#MWUs* (fifth and sixth columns), *#MWUs* continues to exert a facilitatory effect; but importantly, *#ctxt* now has an inhibitory effect on AoFP. This pattern suggests that the initial facilitatory effect of *#ctxt* is due to collinearity with *#MWUs*.

Our results imply that it is involvement in a large number of MWUs – not co-occurrence with a large number of context words – which drives word learning. Furthermore, the effect of MWUs may be limited to MWUs consisting of relatively few words. Hence, when factoring out *#MWUs*, co-occurrence with a large number of context words inhibits acquisition of the target words; and when factoring out the effect of context words, the positive effect of *#MWUs* persists.

Conclusions and Future Work

We began this paper with a review of studies which suggest that children acquire representations of MWUs and that MWUs could facilitate the acquisition of smaller linguistic units contained within them. Based on this, we proposed the *MWU acquisition hypothesis*, according to which the formation of MWU representations precedes and facilitates the formation of individual word representations. The facilitatory effect of *#MWUs* on AoFP supports this hypothesis. More broadly, it supports accounts of language development wherein children acquire linguistic units at various levels of granularity, transitioning gradually from MWUs to more small-grained units.

Our results also have implications for a previous finding: Hills et al. (2010) found that the sum of unique context words occurring within a window of five words to the left and right of each target word predicts age of acquisition of the targets. We also observed a facilitatory effect of *#ctxt*. However, an inhibitory effect of *#ctxt* emerged once *#MWUs* was controlled for. Thus, given that their measure is similar to *#ctxt*, it is possible that Hills et al. (2010)'s result is due to collinearity with the number of MWUs within which target words appear.

In formulating the hypothesis, we purposefully remained agnostic with respect to the specific mechanisms involved in the facilitatory interaction between the acquisition of MWU and single word representations. Accordingly, our results support a general class of theories wherein MWUs are acquired before single words. These could be usage based approaches to language acquisition (Tomasello, 2009), but also proposals such as Peters' (1983), according to which early-acquired MWUs are undersegmented chunks which are gradually segmented into smaller units – units which are themselves stored in memory, where they are again subject to segmentation. In future work, we plan to experiment with differ-

Data set and corpus	Effect (ΔR^2 in %) Covariates baseline	Log-#ctxt	Log-#MWUs	Log-#ctxt unique	Log-#MWUs unique
CBL					
CDS-BE	44.85 ***	1.23 ***	1.73 ***	0.34 (I) ***	0.85 ***
CDS-NA	38.33 ***	0.87 ***	1.35 ***	0.13 (I) ***	0.61 ***
PBS					
CDS-BE	44.85 ***	0.78 ***	1.52 ***	0.55 (I) ***	1.29 ***
CDS-NA	38.33 ***	0.47 ***	1.09 ***	0.18 (I) ***	0.79 ***

Table 5: Effects of log-transformed *#ctxt* and log-transformed *#MWUs*. The effects of *#ctxt* and *#MWUs* were calculated after those of the co-variates had been included. Unique effects are those with the indicated variable entered last (i.e. *#ctxt* after covariates + *#MWUs*, or *#MWUs* after *#ctxt* + covariates). I = inhibitory effect of indicated variable.

ent operationalizations of MWUs, in order to examine what types of MWUs have the strongest potential effect on word learning. This, in turn, may allow us to more closely specify the mechanisms whereby MWUs facilitate word learning.

Acknowledgments

This research was supported by a BOF/TOP grant (ID 29072) of the Research Council of the University of Antwerp.

References

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(02), 239–273.

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth–children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129.

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.

Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: the effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241–248.

Behrens, H. (2009). Usage-based and emergentist approaches to language acquisition. *Linguistics*, *47*(2), 383–411.

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

Brooke, J., Hammond, A., Jacob, D., Tsang, V., Hirst, G., & Shein, F. (2015). Building a lexicon of formulaic language for language learners. In *Proceedings of the 11th workshop on multiword expressions* (pp. 96–104). Denver, Colorado: Association for Computational Linguistics.

Brooke, J., Tsang, V., Hirst, G., & Shein, F. (2014). Unsupervised multiword segmentation of large corpora using prediction-driven decomposition of n-grams. In *Proceedings of coling 2014, the 25th international conference on computational linguistics: Technical papers* (pp. 753–761). Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge: Cambridge University Press.

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*(3), 259–273.

MacWhinney, B. (2000). *The childes project: tools for analyzing talk*. Hillsdale, NJ: Lawrence Erlbaum Associates.

McCauley, S. M., & Christiansen, M. H. (2011). Learning simple statistics for language comprehension and production: The cappuccino model. In L. A. Carlson, C. Hölscher, & T. F. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 1619–24). Austin, TX: Cognitive Science Socitety.

McCauley, S. M., & Christiansen, M. H. (2014). Acquiring formulaic language: A computational model. *The Mental Lexicon*, *9*(3), 419–436.

O'Donnell, T. J. (2015). *Productivity and reuse in language: A theory of linguistic computation and storage*. MIT Press.

Parker, M. D., & Brorson, K. (2005). A comparative study between mean length of utterance in morphemes (mlum) and mean length of utterance in words (mluw). *First Language*, *25*(3), 365–376.

Peters, A. M. (1983). *The units of language acquisition*. Harvard university press.

Tomasello, M. (2009). *Constructing a language*. Harvard university press.

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