TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 9 (2017) 621–636

ISSN:1756-8757 print / 1756-8765 online
DOI: 10.1111/tops.12271

This article is part of the topic "More Than Words: The Role of Multiword Sequences in Language Learning and Use," Morten Christiansen and Inbal Arnon (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/doi/10.1111/tops.2017.9.issue-2/issuetoc

The Role of Multiword Building Blocks in Explaining L1–L2 Differences

Inbal Arnon,[a] Morten H. Christiansen[b]

[a]*Psychology Department, Hebrew University*
[b]*Department of Psychology, Cornell University and School of Communication and Culture, Aarhus University*

Received 1 October 2015; received in revised form 30 August 2016; accepted 7 October 2016

Abstract

Why are children better language learners than adults despite being worse at a range of other cognitive tasks? Here, we explore the role of multiword sequences in explaining L1–L2 differences in learning. In particular, we propose that children and adults differ in their reliance on such multiword units (MWUs) in learning, and that this difference affects learning strategies and outcomes, and leads to difficulty in learning certain grammatical relations. In the first part, we review recent findings that suggest that MWUs play a facilitative role in learning. We then discuss the implications of these findings for L1–L2 differences: We hypothesize that adults are both less likely to extract MWUs and less capable of benefiting from them in the process of learning. In the next section, we draw on psycholinguistic, developmental, and computational findings to support these predictions. We end with a discussion of the relation between this proposal and other accounts of L1–L2 difficulty.

Keywords: First language learning; Second language learning; Multiword units; Chunking; Usage-based models

1. Introduction

Traditionally, words are seen as the basic building blocks of language learning and processing (e.g., Pinker, 1991). Recent years, however, have seen increasing theoretical emphasis on, and empirical evidence for, the idea that multiword sequences, like words,

Correspondence should be sent to Inbal Arnon, Psychology Department, Hebrew University, Mount Scopus, 91905, Jerusalem, Israel. E-mail: inbal.arnon@mail.huji.ac.il

are integral building blocks for language. This idea is found in linguistic approaches that emphasize the role of constructions in language (Culicover & Jackendoff, 2005; Goldberg, 2006) and is advocated in single-system models of language which posit that all linguistic material—whether it is words or larger sequences—is processed by the same cognitive mechanisms (Bybee, 1998; Christiansen & Chater, 2016; Elman, 2009; McClelland, 2010). Speakers are predicted to be sensitive to the properties of both words and multiword patterns. From a developmental perspective, multiword sequences are flagged as important building blocks for language learning in usage-based approaches—where grammatical knowledge is learned by abstracting over stored exemplars of varying sizes and levels of abstraction (e.g., Bannard & Lieven, 2012; Peters, 1983; Theakston & Lieven; Tomasello, 2003). In these approaches such multiword units (MWUs) provide children with lexically specific chunks to be used in early production and allow them to discover distributional and structural relations that hold between words (e.g., Abbot-Smith & Tomasello, 2006). Children are predicted to make use of both words and MWUs in the learning process.

Consistent with this perspective, there is mounting evidence that both children and adults are sensitive to the distributional properties of MWUs and draw on such information in production, comprehension, and learning. Speakers are faster to recognize higher frequency four-word phrases even when all part frequencies are controlled for (e.g., *don't have to worry* vs. *don't have to wait*; Arnon & Snider, 2010; Tremblay, Derwing, Libben, & Westbury, 2011). Strikingly, three-word phrases are responded to just as fast as frequency-matched three-word idioms (Jolsvai, McCauley, & Christiansen, 2013). Speakers show better memory for higher frequency phrases (Tremblay et al., 2011), are able to estimate frequency (Shaoul, Westbury, & Baayen, 2013), and are affected by MWU frequency when processing complex sentences (Reali & Christiansen, 2007). Similar patterns are found in production: Speech onset latencies are affected by MWU frequency (Ellis, Simpson-Vlach, & Maynard, 2008; Janssen & Barber, 2012), as is actual articulation (Arnon & Cohen Priva, 2013; Bybee & Schiebman, 1999). This sensitivity is apparent early on: 2- and 3-year-olds are faster and more accurate in producing higher frequency phrases (e.g., *sit in your chair* vs. *sit in your truck*; Bannard & Matthews, 2008), while 4-year-olds show better production of irregular plurals inside more frequent frames (*Brush your—teeth*; Arnon & Clark, 2011). Taken together, this body of literature highlights the parallels between words and larger phrases, illustrates the way MWUs impact language use in both children and adults, and supports the notion that multiword sequences are integral building blocks in language.

The role of MWUs in language has also been extensively studied in the context of second language (L2) learning (e.g., Schmitt, 2004; Wray, 2004). Actual language use is highly formulaic and restricted, containing a considerable amount of multiword expressions: word combinations that co-occur in predictable and restricted ways (like *slightly different* compared to *mildly different*, see Bannard & Lieven, 2012; Wray, 2002, this volume for reviews). Large-scale corpus studies show that a high proportion (up to 50%) of the language produced by native speakers is formulaic in both written and spoken forms (e.g., De Cock, Granger, Leech, & Mcenery, 1998), a pattern that is found across

languages (see Conklin & Schmitt, 2012 for a review). The ability to use formulaic language is a marker of native fluency and one of the aspects of language that L2 learners struggle with (De Cock et al., 1998; Pawley & Syder, 1980), and that distinguish them from first language (L1) learners (Wray, 2002). Learning how to combine words in native-like ways is one of the biggest challenges for L2 learners (Wray, 1999, 2000, 2004), and even advanced learners produce fewer formulaic sequences than native speakers in both spoken and written production (Paquot & Granger, 2012). Even when L2 learners use formulaic language they seem to do so in a non-native fashion: overusing a limited set of expressions while under-using others (De Cock et al., 1998; Durant & Schmitt, 2009), and not showing the same processing advantages as native speakers (Conklin & Schmitt, 2012). This literature highlights the difficulty that L2 learners have with learning MWUs in a second language and suggests they use them differently from native speakers.

Here, we go beyond existing findings to explore the role of MWUs in explaining L1–L2 differences in learning. While many accounts emphasize the role of multiword sequences in language, and illustrate the relative difficulty that L2 learners have with them, there has not been, to our knowledge, an account that postulates a direct link between the lesser reliance of L2 learners on MWUs and the particular pattern of difficulty they experience in a second language. In particular, we propose that children and adult L2 learners differ in their reliance on MWUs in learning, and that this difference affects learning strategies and outcomes (Arnon, 2010) and leads to difficulty in learning certain grammatical relations. Using MWUs in the learning process can assist in learning semantically opaque relations that hold between consecutive words (like those manifested in grammatical agreement patterns or verb–preposition pairings) that are often difficult for L2 learners to master (e.g., De Keyser, 2005).

In this paper, we first outline two complementary ways in which MWUs emerge during language acquisition, and how they might impact learning. We then substantiate our main hypotheses by reviewing recent findings indicating that MWUs play a facilitative role in learning (Arnon & Ramscar, 2012; Siegelman & Arnon, 2015). This sets the stage for a discussion of the implications for L1–L2differences: We hypothesize that adults are both less likely to extract MWUs and less capable of benefiting from them in the process of learning. Subsequently, we draw on psycholinguistic, developmental, and computational findings to support these predictions. We end with a discussion of the relation between this proposal and other accounts of L1–L2 difficulty.

2. The emergence of multiword building blocks

Before examining the empirical findings on the use of multiword sequences by L1 and L2 speakers, we first consider how speakers in general may come to form such units. We propose two complementary mechanisms: MWUs may be a product of either *undersegmentation*, where a multiword sequence is first acquired as a whole and only later properly segmented, or *chunking*, where patterns of usage lead to the creation of MWUs.

Both processes may work differently in L1 and L2 learning. In our approach, both kinds of MWUs are not stored holistically; rather, the individual words remain activated, with the phrase activating the words and vice versa (similar to the representation of morphologically inflected words; e.g., Baayen, Dijkstra, & Schreuder, 1997). That is, speakers can access the entire sequence in production and comprehension, while still activating the parts (we return to this in the General Discussion).

2.1. Multiword units from undersegmentation

Two factors make undersegmentation less likely in adults: their existing knowledge of words and the input they are exposed to. Unlike adults, infants—who do not know where word boundaries are (or even that words exist)—may extract linguistic units on the basis of prosodic, not lexical information. This claim is consistent with the findings that infants gradually develop the ability to detect smaller prosodic units in fluent speech. For instance, they are capable of detecting utterance-level boundaries before word-level ones (e.g., Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003). Given the nature of child-directed speech, extracting units that correspond to major prosodic breaks will result in a mix of single words, short phrases, and MWUs (Monaghan & Christiansen, 2010). Children's input may further facilitate the extraction of MWUs since it contains many repeated sentence fragments (e.g., *over here*; Fisher & Tokura, 1996) and frames (Cameron-Faulkner, Lieven, & Tomasello, 2003). Importantly, we are not suggesting that children do not segment speech (numerous findings demonstrate infants ability to do so from early on; e.g., Juscyzk & Hohne, 1997), but rather, that during the process of segmentation some of the units they extract will contain more than one lexical word, and that those undersegmented MWUs may co-exist alongside the individual words from which they are made up. These MWUs are still not fully holistic because learners may have partial knowledge of the parts (one of the words, for instance).

The idea that infants' speech segmentation may lead them to extract MWUs (alongside individual words) receives some indirect support from computational models of segmentation. Several models—employing different segmentation strategies—end up with some lexical units that consist of more than one word as a by-product of the cues used to segment speech. The PUDDLE model (Monaghan & Christiansen, 2010) learns to segment child-directed speech by using utterance boundaries to incrementally determine word boundaries in continuous speech (a strategy similar to the one proposed for infants). The model ends up classifying frequently occurring multiword sequences as one "word." Frequent multiword sequences are often classified as one word in computational models of segmentation. For instance, when using transitional probabilities as a cue to word boundary, 30% of the proposed "words" were well-formed multiword sequences (Goldwater, Griffiths, & Johnson, 2009; see also Swingley, 2005). Thus, across a number of recent computational models, MWUs reliably show up as part of the segmentation process.

In contrast to children, there are several reasons to suggest that adult learners will be less likely to undersegment a second language into MWUs. For starters, adults already know that words exist; they have a fundamentally different knowledge of the units of

language than infants do. They are also more capable of directing attention to meaning-carrying units like noun labels or verbs. Their input in itself contains longer sentences and is less repetitive than that of infants (Fisher & Tokura, 1996). Another factor that may lead adults to segment the input into individual words is literacy, which has been found to impacts speakers' awareness of words and decrease the tendency to segment speech into MWUs (Kurvers & Uri, 2006).

2.2. Multiword units from chunking

Repeated exposures to sequences of words that repeatedly co-occur together may lead to multiword representations, for instance, when the combination appears more often than would be expected given the frequency of the parts. Such recurring multiword sequences may form an MWU through chunking processes known from the classic memory literature on serial recall and recognition (see Gobet et al., 2001, for a review). The resulting chunked MWUs exist alongside the words they are made up of. For example, for the MWU *don't have to worry*, speakers are hypothesized to represent information about the entire phrase (e.g., frequency, age of acquisition) as well as about the individual words it is composed of (see Snider & Arnon, 2013 for more details). This information cannot be deduced simply from the properties of the individual words (e.g., the frequency of the phrase *don't have to worry* is not fully predicted from the frequency of the individual words or bigrams). In our approach, such MWUs are not stored holistically; rather, the individual words remain activated, with the phrase activating the words and vice versa.[1] That is, speakers can access the entire sequence in production and comprehension, while still activating the parts (see Snider & Arnon, 2013, for more details).

The difference here between children and adults may lie not in the formation of MWUs (there is little reason to think that chunking as a mechanism works differently in the two populations), but in what is learned from them. Unlike adults, children learning language are, at the same time, learning about the world and how it works. They are learning concepts and words simultaneously (e.g., the word *ball* and the concept it applies to). In contrast, an L2 learner already has both linguistic and conceptual representations in place (from their L1). Although this makes the mapping process easier for individual words (easier to learn the word for *ball* when you already know what a ball is), it also means that L2 learners may miss out on potential predictive relationships that are established when emerging conceptual representations are associated with MWUs. Put differently, adults' existing conceptual knowledge makes them less likely to incorporate grammatical elements into the conceptual representations and more likely to simply map novel labels directly onto existing concepts, making grammatical elements less predictive in the process.

2.3. The impact of multiword units on learning

Our hypothesis that children and adult may learn differently from MWUs is based on two observations. The first is that prediction plays a key role in language: Learning and

processing involve forming predictions about how language unfolds over time (e.g., Altmann & Mirkovic, 2009; Christiansen & Chater, 2016; Levy, 2008). This process is affected by what is already known: Existing knowledge impacts the relative informativity and predictiveness of various cues, which in turn impacts what can be learned from them (e.g., Ramscar, Yarlett, Dye, Denny, & Thorpe, 2010). The second observation is that learning about various aspects of language and the world simultaneously can fundamentally change the informativity of linguistic elements in both undersegmented and chunked MWUs.

Consider, for example, three hypothesized routes to learning article–noun agreement in languages with grammatical gender (see Fig. 1). In the first route, the undersegmented MWU "la + pelota" (the ball) is initially mapped onto the event of experiencing a single ball in the world. Subsequent exposure to the same noun with other articles (e.g., *una* + *pelota*) and to other nouns with the same article (e.g., *la* + *cama*) will eventually lead to the formation of separate noun and article representations while maintaining the early three-way association between the article, the noun, and the concept. Learning segmentation and meaning at the same time makes the article more predictive of both the noun and the concept. In the second route, the novel concept PELOTA will be mapped onto the chunked article–noun sequence "la – pelota" (grouped together based on co-occurrence patterns). Learning concepts and words simultaneously (what PELOTA is and what it is called) will make adjacent words become part of the representation of the developing concepts. That is, the emerging concept of PELOTA will include a predictive relationship with *la*, implicitly encoding the grammatical gender into the noun representation itself. This idea is supported by recent developmental work indicating that labels may function as a kind of additional perceptual feature in category learning and thus become integrated into the acquired concept (e.g., Sloutsky & Fisher, 2012). In the third

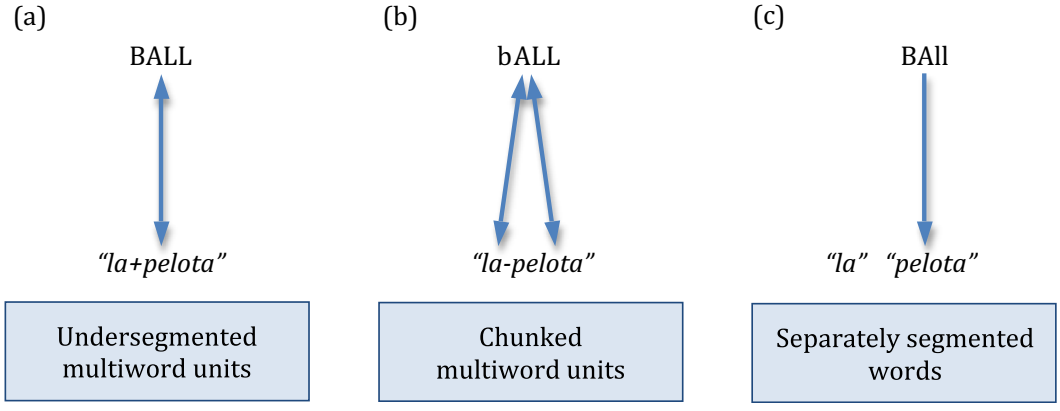

Fig. 1. Illustration of three routes for learning article–noun agreement. (a) The undersegmented multiword unit (MWU) "la + pelota" is mapped onto the emerging concept of BALL, (b) the chunked MWU "la – pelota" is mapped onto emerging BALL concept, and (c) the separately segmented word "pelota" is mapped onto the already existing concept of BALL.

route, two distinguishable words (the article and the noun) are mapped onto already existing conceptual representations (e.g., *pelota* → BALL, *la* → DEFINITE ARTICLE for an English speaker learning Spanish).

The crucial difference between the first two paths and the third lies in the informativity of the grammatical elements (Ramscar et al., 2010). In the two multiword paths, the entire sequence is associated with meaning, making the grammatical element (the article, *la*, in our example) more predictive of both the word following it (the noun *pelota*) and the concept (BALL), even after the independent form and meaning of each element becomes clearer (Ramscar et al., 2010). In contrast, in the individual word path, the initial association between the noun and the object and the fact that the concept is not novel reduces the informativity of the article. Learning from MWUs is expected to enhance learning by increasing the informativity of the grammatical element. This learning advantage can be formally captured in several modeling approaches, including discriminative learning models (e.g., Arnon & Ramscar, 2012; Ellis, 2006; Ramscar et al., 2010).

3. When bigger is better: The advantage of learning from multiword units

Our perspective on how children and adults may benefit from MWUs makes two concrete predictions that have been supported in recent work. The first is that changes in building blocks can result in changes in learning outcomes. Manipulating adult learners' input in a way that directs them to MWUs should enhance performance in certain linguistic domains (but not others). The second is that the benefit of learning from MWUs will be affected by the information conveyed by each of the words: It will be reduced when there is a stronger (semantic) cue available to map the grammatical element to the object.

These predictions have been tested in a series of studies looking at adult learning of grammatical gender and classifier systems in an artificial language. Grammatical gender was chosen because it is a domain that non-native speakers have considerable difficulty with even after extensive exposure (e.g., Scherag, Demuth, Roesler, Neville, & Roeder, 2004). In the first study (Arnon & Ramscar, 2012), participants who were exposed first to a block of full sentences (including the article–noun sequence) and then to a block of single nouns showed better learning of the article–noun pairing compared to participants who were exposed first to nouns and only then to full sentences. Even though both groups had received exactly the same input, only in different orders, participants in the sentence-first condition were better at selecting the correct article in a forced-choice task, and better at producing the correct article for a given noun. Similar results were obtained in simulated acquisition of the language using a Rescorla and Wagner (1972) model: Despite equal exposure to the article–noun pairings, learning was better when exposed first to MWUs.

A second study expanded these findings in several ways (Siegelman & Arnon, 2015). The first experiment established a more direct link between building blocks and learning outcomes by using typing trials to assess the units participants used during learning:

Individually, learners who were more likely to treat the article and noun as one unit showed better learning of the article-noun pairing. A second experiment showed that the facilitative effect of learning from larger units disappears when the article carries semantic information (animate vs. inanimate), illustrating the way unit size interacts with informativity during learning. An additional study expanded this paradigm to natural language (Paul & Gruter, 2016): Monolingual English speakers showed better learning of Chinese classifier–noun associations when exposed first to full sentences and only then to individual vocabulary items. Together, the findings show that learning from MWUs can facilitate adult learning of certain grammatical relations, but not others.

4. The differential role of multiword units in L1 and L2 learning?

These findings—which were obtained from adult learners—may illuminate key differences in L1–L2 learning. Unlike young children, adults rarely reach native proficiency when learning language (a least not across all aspects of the L2). This creates a puzzle: Why are children better language learners than adults despite being worse at a range of other cognitive tasks? Importantly, though, while adult learners clearly experience problems in many linguistic domains—among them pronunciation, morphological, and syntactic processing (Clahsen & Felser, 2006; Johnson & Newport, 1989) and the use of formulaic language (Wray, 2002)—they do not find all aspects of the novel language equally hard and are not always disadvantaged compared to children when it comes to learning a second language. Older learners, for instance, are generally faster and more efficient in the early stages of learning and seem to master certain domains (e.g., vocabulary) better than children (Snow & Hoefnagel-Höhle, 1978). Yet, while some facets of language are learned with relative ease (e.g., vocabulary, word order, yes/no questions, Weber-Fox & Neville, 1996), other aspects—such as grammatical gender, article use, and classifiers—continue to pose difficulty even for highly proficient speakers (DeKeysar, 2005; Johnson & Newport, 1989), including ones classified as near-native (Birdsong, 1992). Critical and sensitive period accounts can predict the qualitative difference in proficiency between the two populations, but they struggle with explaining how the characteristics of the learner lead to the particular pattern of learning difficulties and outcomes we see in children and adults.

Examining the role of MWUs in learning may offer an additional explanation for (some of) the differences between L1 and L2 learning. Specifically, adult L2 learners may be less likely than children to (a) extract undersegmented MWUs and (b) benefit from chunked MWUs in learning because of their existing conceptual knowledge. This could explain adults' difficulty in learning a set of grammatical relations that are consistent, hold between adjacent words, and are semantically non-transparent in that the grammatical element does not make an independent semantic contribution (e.g., gender-marked articles). Language presents us with such relations in the form of dependencies that hold between adjacent words like gender agreement patterns, classifier systems, or verb–preposition pairings, all of which seem difficult for L2 learners to master (see

Clahsen & Felser, 2006; DeKeysar, 2005, for reviews highlighting such aspects as challenging for L2 learners). In contrast, the tendency to learn from single-word building blocks should prove useful for learning vocabulary items, as well as making generalizations that are based on word units, such as word order. Interestingly, adult learners seem to have less difficulty in those domains (Birdsong, 2009; Johnson & Newport, 1989).

The idea that part of adults' difficulty in L2 learning is related to their decreased use of MWUs in learning is based on several assumptions. The first is that children are more likely to form MWUs than adults learning a second language: Larger units are predicted to be more frequent in the inventory of L1 learners compared to adult L2 learners. The second is that children and adults learn differently from multiword sequences: Even when formed, MWUs may be less beneficial for L2 learners, because of their existing linguistic and conceptual knowledge. While the artificial language learning studies reviewed above provide support for the general learning advantage of MWUs, they do not show that such units play a differential role in L1 and L2. Such evidence is, in fact, not straightforward to obtain. It is extremely hard to identify the units of learning in children and adults. By the time they start to talk, children have done a lot of learning already, meaning that there may be a discrepancy between what they produce and what they learned from (Clark & Hecht, 1983). For example, children show a preference for sentences with grammatical forms even when such morphemes are omitted in their own speech (e.g., Shi, Werker, & Cutler, 2006). Similar discrepancies are found in the speech of L2 learners with receptive knowledge typically being more advanced than productive abilities (e.g., Flynn, 1986). How, then, can we see if children and adults differ in their use of multiword building blocks? In the next section, we approach this challenge by drawing on existing developmental, psycholinguistic, and computational findings that together provide preliminary support for these predictions.

5. The role of multiword units in L1 and L2: Developmental, psycholinguistic, and computational evidence

We hypothesized that children and adults differ in the amount of MWUs they draw on and in what they learn from them, but what is the evidence supporting this claim? The first piece of evidence comes from studies of the spontaneous speech of children and adults L2 learners. Children's use of MWUs is supported by findings from their early production and comprehension. Analyses of early child language show that children make use of multiword chunks in early production. Children produce "frozen" multiword utterances at a stage where most of their other productions consist of single words (e.g., Peters, 1983). Many later productions are still not fully productive. Up to 50% of children's early multiword utterances include "frozen" chunks: sequences whose elements are not used productively (Lieven, Behrens, Speares, & Tomasello, 2003). Similar patterns are seen in computational simulations, with many of the primitive units identified in the early stages of language production (1;6–3;0) being multiword sequences (Borensztajn, Zuidema, & Bod, 2009).

Children's error patterns also reflect a reliance on MWUs, which can be traced to frequently occurring sequences in their input. For instance, children's *me*-for-*I* errors (pronoun case errors such as *me do it* where the accusative-marked pronoun is used instead of a nominative one) are related to the proportion of correct preverbal uses (e.g., *let me do it*) in their input (Kirjavainen, Theakston, & Lieven, 2009). In a similar vein, knowledge of multiword patterns can "protect" from errors: Children are less likely to make inversion errors in questions for strings that appeared inverted frequently in the input (e.g., What can X; e.g., Rowland, 2007). Thus, over a range of constructions, children's correct and incorrect productions show evidence of sensitivity to multiword sequences.

Unlike the reports of "chunked" production in child language, the language of L2 learners is often characterized as non-formulaic and is fraught with non-native and collocations (e.g., *put more attention to*, *watch up*; Wray, 2002, 2004; Conklin & Schmitt, 2012). Non-native speakers are also less likely than native speakers to use multiword verbs (e.g., *work out at the gym* vs. *train at the gym*; Siyanova & Schmitt, 2007). It is, moreover, possible that MWUs from a learner's L1 might interfere with acquisition of L2 MWUs[2] (e.g., the English MWU *on the bus* corresponds to the equivalent of "in the bus" in Danish, which could result in interference between L1 and L2 MWUs with a similar meaning). While adult learners in immersion settings clearly learn some fixed expressions early on, they do not seem to use them as building blocks to further grammatical development (Wray, 1999, 2000). Formulaic expressions are even less prominent in the speech of adults learning in a classroom (Conklin & Schmitt, 2012; Wray, 2000, but see Ellis, 2012, who ascribes this difference to frequency of exposure). Additionally, adult L2 learners seem to treat sequences of language as more flexible than they actually are (e.g., changing word order in collocations; Wray, 2004).

Taken together, these findings suggest that L2 learners differ from native speakers in their ability to track and make use of MWUs. However, the data from spontaneous speech are incomplete: the data do not demonstrate children's use of larger units during learning (rather than just during production), and they do not provide a direct comparison of the amount of MWUs in the speech of L1 and L2 learners (it looks at the two groups of learners separately). We draw on two additional sources of evidence in support of these claims. The first shows that MWUs serve as building blocks in L1 learning by drawing on lexical Age-of-Acquisition (AoA) effects. The second uses a series of computational studies to directly test the prediction that adult L2 learners' speech is less "chunky" than that of child learner and adult native speakers.

Words that are acquired earlier in childhood show processing advantages in adults on a variety of tasks (see Juhasz, 2005, for review). These lexical AoA effects illustrate the lasting impact of early-learned material on adult language and offer a way of examining early units by looking at adult processing. If MWUs also serve as building blocks for language, they should exhibit AoA effects. Recent work has tested this prediction in two reaction time studies with adults (Arnon, McCauley, & Christiansen, 2017). Early-acquired phrases were identified using a combination of corpus-based measures and subjective ratings. In two separate studies adults responded faster to early-acquired three-word phrases (estimated using corpus frequencies and subjective ratings) compared to

later-acquired ones, after controlling for all adult frequencies, lexical AoA, and plausibility (e.g., faster to *take them off* compared to *take time off*). These findings provide support for the role of MWUs in L1 learning: Like early-acquired words, early-acquired multiword sequences maintain a privileged status in adult processing.

Preliminary support for a different use of MWUs in adult L2 learning comes from McCauley and Christiansen (this volume), who used the chunk-based learner (CBL) model (McCauley & Christiansen, 2014) as a computational implementation of the Traceback method (Lieven et al., 2003) to compare the "chunkedness" of L1 and L2 productions. CBL learns by chunking words together based on statistical information to form units used in language comprehension and production. The model has previously been used to model data on children's use of multiword chunks. Using matched corpora of L1 and L2 learner speech as input to CBL, the model compared the chunkedness of language produced by adult L2 learners (native Italian speakers learning English or German) to that of child and adult native speakers of the same languages (English and German). The basic question was will the speech of native speakers contain more multiword chunks than that of adult L2 learners? The results showed that, compared to those of the L2 speakers, the productions of the native speakers—whether children or adults—were considerably more chunked as measured by repeated multiword sequences. These findings support the idea that adult L2 learners have a harder time chunking incoming input in a second language, a factor that is likely to negatively impact their mastery of grammatical regularities such as morphology and grammatical gender.

6. General discussion

In this paper, we have proposed that adults are less likely to draw on MWUs in the process of learning a second language, and that this leads to particular patterns of difficulty. This view naturally addresses two challenges that face theories of L2 learning: the challenge of *how* to link specific features of the learner to the learning outcomes observed and the challenge of *why* L2 learners find certain linguistic domains harder than others. Adult learners will be less likely to form MWUs (because of their prior knowledge about words, the input they are exposed to, and their literacy) and less likely to use them during conceptual learning (because of what they already know about words and concepts). This, in turn, will make it harder to learn certain grammatical relations (the *how* question). This difficulty will be especially pronounced when the relations in question involve more than one word and are more semantically opaque; in such cases, the information carried by each element on its own is not enough to consolidate the link between them (the *why* question).

Our proposal diverges from existing accounts of L1–L2 differences in several ways. We embed the question within the more general issue of language representation, turning to models of first language learning for inspiration. Unlike critical or sensitive period approaches, we emphasize the way experience (rather than maturational or cognitive differences) shapes the building blocks learners utilize, changing their informativity (and

learnability) in the process. The resulting account is gradient: Learning outcomes will differ according to what needs to be learned, allowing for differential performance on varying aspects of L2; and multidimensional: success will be affected not only by age but by a host of factors impacting the building blocks used, including the learning situation (e.g., explicit vs. implicit), modality of presentation (written vs. spoken), and the characteristics of the L1 (see also Hakuta, Bialystok, & Wiley, 2003; Munoz & Singleton, 2011).

Of course, emphasizing the role of MWUs in learning cannot explain all the differences between L1 and L2 learning. The path of L2 learning, as much research has shown, is highly variable and is impacted by the characteristics of the first and second languages, the environment in which learning takes place, and the cognitive and social characteristics of the learner (see Dornyei & Skehan, 2003; Ellis & Shintani, 2013, for overview). However, we believe that looking at the role of MWUs in language in general, and their possibly different role in L1 and L2 learning, can provide a roadmap for establishing links between learner characteristics and learning outcomes and predicting which aspects of a novel language adults will find particularly challenging. Much research has yet to be done to test the link between building blocks and learning outcomes, and to identify the building blocks used by the different types of learners. In particular, while there is evidence that features can become part of a novel concept representation (Sloutsky & Fisher, 2012), there is no work that examines the effect of multiword building blocks on conceptual learning. That is, there is little evidence so far for the claim that simultaneously learning word sequences and developing the concepts to which they refer can benefit learning.

We have made several claims about the way MWUs are learned and represented. First, the existence of undersegmented MWUs suggests that speakers store multiword sequences. While such sequences do not have to be holistic (without any word representation), the prediction is that larger patterns are stored and may be learned before the representation of the parts. Second, the existence of chunked MWUs is consistent with a redundant view of the lexicon (e.g., Elman, 2009) where both parts and wholes are simultaneously represented. Both claims are inspired by the morphological literature on the representation of morphologically inflected words (e.g., Baayen et al., 1997). However, there are other ways to think about how multiword patterns come out. In particular, certain discriminative learning models challenge both of these claims and suggest instead that that the sensitivity to multiword information can emerge without having any word or MWU representations (Baayen, Hendrick, & Ramscar, 2013; Ramscar et al., 2010). Multiword frequency effects have been captured using only letter representations in such models, though it remains to be seen whether they can capture the full range of findings on MWUs. For example, it is unclear how such models would capture effects of MWUs on speech duration in language production (e.g., Arnon & Cohen Priva, 2013) and AoA effects (Arnon et al., 2017).

In sum, in this paper we built on what are usually disparate lines of research—usage-based studies of first language learning, psycholinguistic investigations of multiword processing, and studies of L2 learning—to explore the role of multiword building blocks in L1–L2 differences. Theoretically, we emphasize the complex interaction between

experience, building blocks, and learning outcomes to suggest a more nuanced perspective on why adults struggle with learning a second language. From an applied perspective, our proposal has the potential to significantly alter how we teach adults a second language. By investigating the way early units affect learning outcomes, and highlighting the role of larger chunks in learning, we may be able to generate novel ways to improve L2 learning, a task of utmost importance in our increasingly multilingual world.

Acknowledgments

We would like to thank Harald Baayen, Adele Goldberg, and several anonymous reviewers for their comments on previous versions of this paper. This work was partially supported by BSF grant number 2011107.

Notes

1. Undersegmented MWUs might gradually become chunked MWUs after segmentation if they occur frequently enough.
2. We thank Adele Goldberg for suggesting this complementary factor that may further exacerbate the problems L2 learners experience with MWUs.

References

Abbot-Smith, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage based account of syntactic acquisition. *The Linguistic Review*, *23*, 275–290.

Altmann, G., & Mirkovic, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, *33*, 583–609.

Arnon, I. (2010). Starting big: The role of multiword phrases in language learning and use. Unpublished PhD dissertation, Stanford University.

Arnon, I., & Clark, E. V. (2011). When '*on your feet*' is better than '*feet*': Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*, 107–129.

Arnon, I., & Cohen Priva, U. (2013). More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*, 349–373.

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, *92*, 265–280.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*, 292–305.

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

Baayen, R. H., Dijkstra, T., & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language*, *37*, 94–117.

Baayen, R. H., Hendrix, P., & Ramscar, M. (2013). Sidestepping the combinatorial explosion: Towards a processing model based on discriminative learning. *Language and Speech*, *56*, 329–349.

Bannard, C., & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics*, *32*, 3–16.

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

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, *68*, 706–755.

Birdsong, D. (2009). Age and the end state of second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.), *The New Handbook of Second Language Acquisition*. Amsterdam: Elsevier.

Borensztajn, G., Zuidema, W., & Bod, R. (2009). Children's grammars grow more abstract with age. *Topics in Cognitive Science*, *1*, 175–188.

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

Bybee, J., & Schiebman, J. (1999). The effect of usage on degrees of constituency: The reduction of don't in English. *Linguistics*, *37*, 575–596.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child-directed speech. *Cognitive Science*, *27*, 843–873.

Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences*, *39*, 62–102.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.

Clark, E. V., & Hecht, B. F. (1983). Comprehension, production, and language acquisition. *Annual Review of Psychology*, *34*, 325–349.

Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45–61.

Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford, UK: Oxford University Press.

De Cock, S., Granger, S., Leech, G., & Mcenery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67–79). London & New York: Addison Wesley Longman.

De Keysar, R. M. (2005). What makes learning second language grammar difficult? A review of issues. *Language Learning*, *55*, 1–25.

Dornyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Malden, MA: Blackwell.

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

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking and perceptual learning. *Applied Linguistics*, *27*, 164–194.

Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, *32*, 17–44.

Ellis, R., & Shintani, N. (2013). *Exploring language pedagogy through second language acquisition research*. London: Routledge.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second-language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, *41*, 375–396.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, *33*, 547–582.

Fisher, C. L., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant directed speech: Cross-linguistic evidence. *Child Development*, *67*, 3192–3218.

Flynn, S. (1986). Production vs. comprehension: Differences in underlying competence. *Studies in Second Language Acquisition*, *8*, 135–164.

Gobet, F., Lane, P. C. R., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*, 236–243.

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

Goldwater, S., Griffiths, T. L., & Johnson, M. (2009). A Bayesian framework for word segmentation: Exploring the effects of context. *Cognition*, *112*, 21–54.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the Critical Period Hypothesis for second language acquisition. *Psychological Science*, *14*, 31–38.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLoS ONE*, *7*, https://doi.org/10.1371/journal.pone.0033202.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60–99.

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2013). Meaning overrides frequency in idiomatic and compositional multiword chunks. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th Annual Conference of the Cognitive Science Society* (pp. 692–697). Austin, TX: Cognitive Science Society.

Juhasz, B. J. (2005). Age-of-Acquisition effects in word and picture identification. *Psychological Bulletin*, *131*, 684–712.

Juscyzk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science*, *277*, 1984–1986.

Kirjavainen, M. M. M., Theakston, A. L., & Lieven, E. V. (2009). Can input explain children's me-for-I errors? *Journal of Child Language*, *36*, 1091–1114.

Kurvers, J., & Uri, H. (2006). Metalexical awareness: Development, methodology or written language? *Journal of Psycholinguistic Research*, *35*, 353–367.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Lieven, E. V., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language*, *30*, 333–370.

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

McClelland, J. L. (2010). Emergence in cognitive science. *Topics in Cognitive Science*, *2*, 751–770.

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*, 545–564.

Munoz, C., & Singleton, D. (2011). A critical review of age-related research on L2 ultimate attainment. *Language and Teaching*, *44*, 1–35.

Paul, J. Z., & Gruter, T. (2016). Blocking effects in the learning of Chinese classifiers. *Language Learning*, *66*, 972–999.

Pawley, A., & Syder, F. H. (1980). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. Schmidt (Eds.), *Communicative competence* (pp. 191–225). London: Longmans.

Peters, A. M. (1983). *The units of language acquisition*. Cambridge, UK: Cambridge University Press.

Pinker, S. (1991). Rules of language. *Science*, *253*, 530–535.

Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K. (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive Science*, *34*, 909–957.

Reali, F., & Christiansen, M. H. (2007). Word-chunk frequencies affect the processing of pronominal object-relative clauses. *Quarterly Journal of Experimental Psychology*, *60*, 161–170.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian Conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rowland, C. F. (2007). Explaining errors in children's questions. *Cognition*, *104*, 106–134.

Scherag, A., Demuth, L., Roesler, F., Neville, H. J., & Roeder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, *93*, B97–B108.

Schmitt, N. (2004). *Formulaic Sequences: Acquisition, processing and use*. Amsterdam: John Benjamins.

Shaoul, C., Westbury, C. F., & Baayen, R. H. (2013). The subjective frequency of word n-grams. *Psihologija*, *46*, 497–537.

Shi, R., Werker, J. F., & Culter, A. (2006). Recognition and Representation of function words. *Infancy*, *10*, 187–198.

Siegelman, N., & Arnon, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Language and Memory*, *85*, 60–75.

Siyanova, A., & Schmitt, N. (2007). Native and non-native use of multi-word vs. one-word verbs. *International Review of Applied Linguistics*, *45*, 119–139.

Sloustky, V. M., & Fisher, A. V. (2012). Linguistic labels: conceptual markers or object features? *Journal of Experimental Child Psychology*, *111*, 65–86.

Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and noncompositional phrases in the lexicon. In S. Gries, & D. Divjak (Eds.), *Frequency effects in language*. Berlin: Mouton de Gruyter.

Snow, C. E., & Hoefnagel-Höhle, E. (1978). The critical period for language acquisition: Evidence from second language learning. *Child Development*, *49*, 1114–1128.

Soderstrom, M., Seidl, A., Kemler Nelson, D. G., & Jusczyk, P. W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, *49*, 249–267.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*, 86–132.

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

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

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specialization for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*, 231–256.

Wray, A. (1999). Formulaic language in learners and native speakers. *Language Teaching*, *32*, 213–231.

Wray, A. (2000). Formulaic sequences in second language teaching: Principles and practice. *Applied Linguistics*, *21*, 463–489.

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

Wray, A. (2004). Here's one I prepared earlier: Formulaic language learning on television. In N. Schmitt (Ed.), *The acquisition and use of formulaic sequences* (pp. 249–268). Amsterdam: John Benjamins.