Chapter 4 Formulaic language and speech prosody

Phoebe Lin

Department of English, The Hong Kong Polytechnic University

Abstract

On the basis that most first language formulaic sequences are acquired and accessed auditorily, this chapter argues for the fundamental importance of the prosodic representation of formulaic sequences in the mental lexicon. It also critically examines the suggestion that prosody can offer reliable cues to facilitate formulaic sequence acquisition and that second language learners' lack of exposure to spoken input may have led to their difficulties with acquiring second language formulaic sequences.

Keywords: formulaic sequences, formulaic language prosody, challenges of L2 vocabulary acquisition, prosodic processing ability, spoken input in L2 vocabulary acquisition

1. Introduction

> The trouble is, organic produce is not cheap.
>
> (British National Corpus, ARJ 325)

The phrases the thing is and the trouble is are curious cases. They are so familiar and frequent in our everyday speech that they require little effort to produce or comprehend. Parts of these phrases can be skipped in fast speech (i.e., saying trouble is and thing is instead) and yet hearers will have no problem with comprehension. These characteristics explain why the phrases are considered examples of formulaic sequences (FSs). However, the peculiarities in the way that these phrases are articulated are frequently overlooked. They are often articulated in one intonation unit and/or are followed by a pause. This is probably the reason why writers regularly insert a comma after the thing is and the trouble is. This is intriguing because intonation unit boundaries and pauses often coincide with phrase or clause boundaries, which means that a prosodic break should be inserted between the subject (i.e., the trouble) and the predicate (i.e., is organic produce is not cheap). The result of this, however, sounds unnatural (i.e., *The trouble, is

organic produce is not cheap). A second peculiarity about this example concerns its accentuation pattern. If a speaker wants to emphasize the meaning of the phrases, the function word will be accented (i.e., the trouble IS) even though the tendency in English is for accent to be assigned to the lexical word (i.e., the TROUble is). These peculiarities have led to the question of whether FSs might be subject to some special prosodic rules.

2. Is there a prosody special to formulaic language?

Beyond the anecdotal observations presented earlier, psycholinguistic evidence seems to point towards the existence of prosodic patterns that may be special to FSs. In their seminal papers, van Lancker and Canter (1981) and van Lancker, Canter, and Terbeek (1981) investigated whether and how prosody might differ when a class of idioms, which they called ditropic sentences, were articulated with either the intended idiomatic or literal meanings. An example of a ditropic sentence is "she was to keep a stiff upper lip", which can mean, idiomatically, that she is brave or, literally, that she contracts her lip muscles. In their first study, van Lancker and Canter (1981) asked five male speakers to read aloud 15 pairs of ditropic sentences embedded in disambiguating paragraphs. Recordings of the ditropic sentences were then excised from context and played to native English-speaking listeners. However, the listeners were unable to identify whether the stimuli carried the intended idiomatic or literal meanings with above chance accuracy. In their second study, two male speakers read aloud the same 15 pairs of ditropic sentences without context and were instructed to convey the pairs' contrasting meanings as distinctly as possible. This time, the listeners were able to judge, with above chance accuracy, the intended meaning of the stimuli. In the follow-up acoustic analyses, van Lancker et al. (1981) found that these (American English) ditropic sentences could be disambiguated because, compared with their counterparts, the readings with intended literal meanings had longer sentence duration, more internal pauses, more juncture, more pitch contours, and were systematically marked by Accent A as defined by Bolinger (1965). This investigation of the acoustic differences between the idiomatic and literal interpretations of ditropic sentences was later extended to French (Abdelli-Baruh, Yang, Ahn, & van Lancker, 2007) and Korean (Yang, Ahn, & van Lancker, 2010). More recently, Siyanova-Chanturia and Lin (2017) have delved deeper into the question by including matched novel control phrases in the acoustic comparisons. In this experiment, 66 native speakers of British English read aloud three types of stimuli (i.e., idioms used figuratively, idioms used literally, and matched novel control phrases) embedded in disambiguating paragraphs. The acoustic analysis revealed that the

idioms had shorter durations than the controls. Furthermore, the idioms also had shorter durations when used figuratively than when used literally.

The prosodic patterns of FSs are also a major area of interest in spoken corpus research. Researchers are particularly keen on detecting patterns in the prosodic features of FSs because such findings will not only shed light on the processing of FSs from a usage-based perspective, but also have substantial practical implications for language teaching and Natural Language Processing (see Lin, 2015, 2018, for further discussion). So far, the list of prosodic features found to be closely associated with FSs includes alignment with intonation unit boundaries (Lin, 2010b; Lin & Adolphs, 2009) and pauses (Wray, 2004), faster rhythm (Lin, 2010b), which may or may not be accompanied by phonological reduction (Bybee & Scheibman, 1999), resistance to internal pauses or hesitations (Wray, 2004), and restricted accentuation patterns (Aijmer, 1996; Lin, 2013). These prosodic patterns, which are opaque for language learners and computers, are vital to the accurate communication of meaning of FSs. To illustrate the point, consider the accentuation pattern of the idiom my ears are burning (Ashby, 2006). The accurate use of the idiom necessitates the assignment of accent to the word burning (i.e., my ears are BURNing). Alternative accentuation patterns (e.g., my EARS are burning) are unacceptable. Similarly, for FSs which contain flexible slots (e.g., as far as ___ is concerned, from the ___ point of view), the default pattern is to accent the word in the flexible slot (Lin, 2013). Deviation from this pattern, though possible, is rare.

3. Rationale and aims of the chapter

This chapter builds upon the aforementioned work describing the prosody of FSs in psycholinguistics and corpus linguistics, and explores the wider cognitive picture within which FSs and speech prosody are situated. This desire to search for a wider cognitive perspective on the relationship between FSs and speech prosody emerged 10 years ago (i.e., Lin & Adolphs, 2009) when we became aware of the fundamental importance of prosody to the contextual meaning of spoken FSs. At that time, corpus linguists tended to apply the same approaches to analyzing FS use in spoken and written corpora. This practice of overlooking prosodic information is questionable because the pragmatic meaning and function of spoken FSs are encoded in and vary with their prosody (see Lin, 2010b for further discussion). On delving deeper into the nature of spoken FSs, the fundamental importance of spoken FSs in everyday communication became even more apparent. As Carter (personal communication, 2012) suggests, the proportion of spoken to written communication in an ordinary person's daily linguistic encounter is believed to be 9:1. If first language (L1)

speakers are nine times more likely to acquire and use FSs in speech than in written communication, then the phonological representation of formulaic language (FL) should be far more frequently activated than, say, its orthographic representation. This overwhelming frequency with which L1 speakers activate the phonological representation of FSs justifies why the phonological form of FSs deserves significant research attention. In fact, given that L1 FSs are acquired and accessed through the auditory channel during extensive exposure to spoken communication, a new perspective on FSs may be necessary in which FSs are reconceptualized (and redefined) as strings of sounds, rather than strings of words.

Although the earliest research into the prosody of FSs dates back to the 1970s (e.g., Bloom, 1973; Peters, 1977), and important breakthroughs were made in the 1980s (e.g., Peters, 1983; van Lancker & Canter, 1981; van Lancker et al., 1981), very few empirical studies on this topic have been conducted since then. Recently, however, there has been a renewed interest in the topic within the field of corpus linguistics (e.g., Lin, 2010a, 2010b, 2013, 2018; Lin & Adolphs, 2009) because of the practical applications that the findings will have for both English language teaching and Natural Language Processing. These corpus studies, however, have stopped at the point of capturing the prosodic patterns of automatically identified word sequences. They did not delve further into the cognitive bases of underlying the systematicity observed in the prosodic patterns.

This chapter will fill this knowledge gap and explore the uncharted waters of FL and speech prosody from a cognitive perspective. For this exploration, an FS is broadly defined as "a sequence, continuous or discontinuous, of words or other elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar" (Wray, 2002, p. 9). As an umbrella term, FL subsumes many types of lexicalized word combinations ranging from idioms to proverbs, clichés, conversational routines, collocations, and so on (see Wray, 2002, for a comprehensive review). This chapter examines the idea of speech prosody as a perceptual category that is the result of interactions between four acoustic components, namely pitch, loudness, timing, and voice quality. In the phonological literature, pitch is considered the main contributor to the perception of tones; loudness is considered the main contributor to the perception of stress; and timing is the main contributor to the perception of pauses and rhythm (see Cruttenden, 1997, for further discussion).

The challenge of offering a critical review of FL and speech prosody from a cognitive perspective is substantial, not least because very few empirical studies have directly addressed this topic. Furthermore,

the notions of FL and speech prosody are both very broad and complex; therefore, controversies still surround what each entails and how they can be quantified. It is with this background in mind that this chapter provides a critical review of the relevant research findings. Although this chapter attempts to bring together research evidence from across several disciplines, such as phonology, child language research, corpus linguistics, psychology, and biology, the discussion may still appear limited because there are, truly, many research gaps on this topic waiting to be addressed. This review aims to highlight the need for more empirical research on the cognitive aspects of FL and speech prosody and, at the same time, identify possible directions for future empirical studies.

4. The emergence of formulaic sequences in child first language acquisition

Since the input that children receive is 100% spoken, the phonological representation of FSs is bound to play a prominent role in the mental lexicon of the child. Consequently, child L1 acquisition probably offers the best example to support the conceptualization of FSs as strings of sounds, rather than strings of words.

When comparing the two levels of phonological representation (i.e., prosody and phonemics), prosody appears to play the more fundamental role in L1 acquisition. In one of the earliest FL studies, Peters (1977) examined the productions of an 11-month-old L1 learner, Minh. In Minh's speech, Peters noticed stretches of vocalizations that displayed a distinctive and adult speech-like melody. Despite the lack of phonemic articulatory precision, these vocalizations could be recognized in their immediate speech context as the child's attempts at producing context-bound utterances such as look at that! and what's that?. That the child was able to produce these utterances was unexpected because he had clearly not yet mastered the grammar, vocabulary, or the phonemes to realize those utterances. Peters (1974) describes this phenomenon as children "learning of a tune before the words". In fact, children's ability to memorize tunes and melodies holistically is remarkable, as people who have spent time with toddlers may observe. Wray (2002) also discussed the case of a 21-month-old child, Ellen, who had managed to memorize and sing the tune Away in a Manger perfectly, even though she understood neither the words nor the meaning of the carol. The original lyrics "Away in a manger, no crib for a bed. The little Lord Jesus lay down his sweet head" came out in Ellen's singing as "Away in a manger, no crisp for a bear. The little Lord Jesus lay down on his hair". (see Wray, 2002, p. 109, for further information). While these mis-sung lyrics may bring a light-hearted smile, they nonetheless also provide evidence for the pre-eminence of prosody in child L1 acquisition. When children are exposed to L1 spoken input, the melody (or, indeed, prosody) is very important because it guides their attention to and segmentation of the spoken input. Utterances which form prosodically

holistic units such as look at that! and what's that? are acquired and produced holistically. The length of these prosodically holistic units can extend to the whole tune, as we can see in the case of the carol Away in a Manger, but they can also be a shorter phrase, as in the case of Look at that!. In both cases, prosody is clearly prioritized; the fact that the child does not understand the meaning or the pronunciation of the words constituting the prosodically holistic units does not appear to hinder the child's acquisition or production of the units. Likewise, the lack of phonemic articulatory precision does not prevent child language researchers from recognizing the presence of holistically learned chunks in the child L1 learner's speech.

That child L1 learners prioritize learning the tune before the words should not be surprising. Recent research has indicated that foetuses and infants are biologically predisposed to rely on prosodic cues to acquire their L1. Due to the abdominal barrier, speech from the outside world is filtered before it reaches the foetus's ears. The vowels and consonants are mumbled, leaving only prosodic cues as the remains of the speech signal (imagine hearing speech under water). For this reason, human foetuses have learned to tune in to the prosodic patterns in the speech signal. Empirical evidence to support the theory that human foetuses tune in and react to the prosodic patterns of their L1 comes from many sources, including a number of studies in biology which recorded and analyzed the cries of newborn babies (e.g., Mampe, Friederici, Christophe, & Wermke, 2009; Wermke & Mende, 2009). In Mampe et al.'s (2009) study, the researchers analyzed the cries of 30 French and 30 German newborn babies and found that the melody of the cries mimicked the prosody of the babies' mother tongues. Since these were newborn babies, it was clear that their acquisition of the melody of their mother tongues had taken place in the womb. This finding has been triangulated by other studies, which have adopted alternative methods to trace the beginning of foetal sensitivity to and memory of linguistic prosody (e.g., DeCasper & Spence, 1986; Ferrari et al., 2016). In DeCasper and Spence's (1986) classic study, pregnant women were asked to read to their foetuses each day during the last 6 weeks of pregnancy. Based on the newborns' response to the sounds of these read passages, the researchers concluded that sensitivity to and memory of linguistic prosody can be traced back to the last trimester of pregnancy. More recently, Ferrari et al. (2016) used two-dimensional ultrasound to directly measure foetal mouth movements in response to their mothers' speech. The study found that, even at 25 weeks of gestation (i.e., the start of the third trimester), foetuses seem to already be responsive to maternal stimuli. These studies involving prenatal and newborn subjects point clearly towards the fact that human foetuses are biologically predisposed to rely on prosodic cues for L1 acquisition, a reliance which seems to extend to infancy (as we see in the cases discussed in Peters, 1977; Bloom, 1973) and toddlerhood (Bannard & Matthews, 2008; Wray, 2002).

5. Prosody, grammar, and formulaic sequence acquisition

Prosody plays a vital role in many aspects of L1 acquisition. In the literature, researchers have long been interested in the ways that prosody facilitates L1 grammar acquisition. For example, several studies have investigated whether prosody can offer reliable cues to syntactic and phase structure (e.g., Jusczyk et al., 1992; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright Cassidy, 1989). In the prosody literature, prosodic breaks have been classified based on their perceived level of discontinuity. Based on Knowles's (1991) classification, pauses accompanied by audible breathing represent the strongest type of prosodic break (level 5), followed by pause (level 4), pitch discontinuity (level 3), segmental separation features (level 2), segmental run-on cancelled (level 1), and nothing measurable (level 0). What the empirical evidence has shown is that the strength of prosodic breaks may offer child L1 learner cues concerning the hierarchical organization of phrases and syntactic units. For instance, studies by Goldman-Eisler (1972) and Scott (1982) found that pauses (levels 4 and 5 prosodic breaks) were more likely to occur at major syntactic boundaries than arbitrarily within phrases. Pause length also appears to reflect the hierarchical organization of phrases (e.g., Cooper & Paccia-Cooper, 1980; Gee & Grosjean, 1983). Before major syntactic boundaries, the final syllables of the syntactic units tend to be lengthened as well as showing a global decline in pitch (e.g., Cruttenden, 1997; Wichmann, 2000).

Other, corpus-based, studies have examined this question concerning prosodic cues to syntactic and phrase structure from a quantitative angle by asking what kind of syntactic unit do intonation units (IUs) most commonly correspond to between different languages (e.g., Iwasaki & Tao, 1993; Schuetze-Coburn, 1994). These studies have consistently found clauses to be the prototypical correlates of intonation units. In Iwasaki and Tao's (1993) study, 53.6%, 45.4%, and 39.8% of the intonation units in their corpora of English, Japanese, and Mandarin Chinese, respectively, were clauses. These figures are interesting because, contrary to popular belief, the correspondence between intonation units and clauses is far lower than 100%. If prosody guides the segmentation of the input speech signal and influences the unit of lexical storage in the mental lexicon, then the unit of lexical storage could differ according to the child's L1. It seems logical to think that the prosody of L1 Mandarin Chinese (which showed 39.8% of IU-clause correspondence) would be more likely to store alternative types of units (other than clauses) in the mental lexicon when compared with the prosody of L1 English (which showed 53.6% of IU-clause correspondence). Previous research (e.g., Schuetze-Coburn, 1994) has provided the example of sentence fragments as possible alternative types of units with which intonation units may correspond. However, Lin's corpus studies (reviewed later)

indicate that intonation units may also correspond to FSs. In other words, following the same logic as L1 grammar acquisition studies (e.g., Jusczyk et al., 1992; Kemler Nelson et al., 1989), it seems that, statistically speaking, prosody is also likely to facilitate the acquisition of L1 FSs.

The fact that adjacent words that form grammatical units appear to be readily chunked by prosodic cues in spoken input (see Iwasaki & Tao's (1993) figures cited earlier) has convinced researchers that prosody serves a bootstrapping function in language acquisition (e.g., Jusczyk et al., 1992; Kemler Nelson et al., 1989). What this means, as Fisher and Tokura (1996) put it, is that prosodic information could provide learners with a linguistically useful, bottom-up means of segmenting speech. Such prosodic evidence, when combined with distributional analyses, may help learners to "rule out conjectures about grammatical structure inconsistent with the audible structure of input sentences" (p. 344).

Since research into the prosodic features of FSs is still in its infancy, there has been extremely little empirical research on the role of prosody in the L1 FS acquisition mechanism. Nevertheless, taking into consideration Lin and her colleagues' corpus research (Lin, 2010b, 2013; Lin & Adolphs, 2009), which found that FSs aligned with pitch discontinuity (level 3 prosodic breaks in Knowles' 1991 classification) approximately 50% of the time, it seems logical to assume that prosody can offer reliable cues which facilitate the acquisition of L1 FSs. In other words, prosody may also serve a bootstrapping function in L1 FS acquisition as well as L1 grammar acquisition.

In psychology and child language acquisition, the closest, indirect evidence for prosody's bootstrapping function in L1 FS acquisition so far comes from Peters (1977, 1983), whose findings seem to illustrate the fact that FSs and prosody are inseparable in the case of child L1 acquisition. The reason for suggesting that Peters' (1977, 1983) studies provide indirect evidence is that these studies' original aim was to draw attention to the fact that language development in children does not appear to proceed in a linear fashion. One-word utterances are not necessarily mastered before two-word utterances or two-word utterances before three-word utterances. Instead, it is also possible for children to memorize and produce larger chunks such as What's that? and Look at that! even before they master the words or the grammatical structures that constitute those chunks. These studies advocate a careful distinction between utterances which the child has acquired and produced holistically, such as What's that? and Look at that!, and utterances which are generated by articulating single words in succession. This is because such a distinction will enable child language researchers to reach a more accurate diagnosis of a child's stage of linguistic development

(Branigan, 1979; Peters, 1983). The key to distinguishing between these two types of utterances lies, above all, in examining their prosody. As mentioned previously, utterances learned holistically appear to be characterized by their distinctive adult speech-like melody. The words constituting the utterances may be pronounced unclearly and imprecisely, but the delivery, nevertheless, appears to be distinctively fluent. Utterances formed by articulating single words in succession, on the other hand, have an "unmistakable" prosodic pattern whereby "[e]ach word occurred with terminal falling pitch contour, and relative equal stress, and there was a variable but distinct pause between them so that utterance boundaries were clearly marked" (Bloom, 1973, p. 41). Given these findings, there seems to be a clear circularity between FSs and prosody in the case of child L1 acquisition (see also Lin, 2010a). On the one hand, it has been suggested that prosody facilitates child L1 acquisition; but, on the other hand, holistically acquired and memorized chunks (i.e., FSs) in child language are identified based on their prosody. This is why FSs and prosody seem inseparable in the case of child L1 acquisition.

As the child L1 learner's linguistic competence develops, these holistically acquired and memorized chunks eventually undergo one of two processes. They may be "unpacked" as their internal structure is analyzed to different extents as part of the L1 grammar acquisition process, or they may remain as holistic units in the mental lexicon (see Wray, 2002, for further discussion). In the absence of direct empirical research, our knowledge of the ways in which the unpacking process may affect the mental prosodic representation as well as the prosodic realization of the analyzed chunks remains limited. That said, with the advent of dense corpora of child-directed speech (e.g., Bannard & Matthews, 2008), necessary data already exist to assist in answering these important questions.

6. The challenges of formulaic sequence acquisition in a second language

Knowledge of FSs is considered key to the achievement of nativelike speech fluency (Pawley & Syder, 1983; Wood, 2012). In the literature, many anecdotes have been presented concerning adult second language (L2) learners who seem to be able to speak with distinctive fluency as soon as they start speaking formulaically (see, e.g., Dechert, 1983; Raupach, 1984). In one case (Wray, 2004), a beginner learner was able to give cooking demonstrations on television in Welsh after spending just five days learning context-specific FSs. The fact that FSs seem to offer a "shortcut" to speech fluency has intrigued ELT (English Language Teaching) professionals. It is believed that L2 learners' speech fluency problems may be partially tackled by facilitating their acquisition of L2 FSs (Meunier & Granger, 2008; Wood, 2012).

Despite the perceived importance of FSs to nativelike fluency, the acquisition of FSs in the L2 does not appear to be straightforward. Studies have reported many challenges confronting L2 FS acquisition. The problems include L2 learners' lack of breadth of FS knowledge, difficulty with memorizing FSs without introducing formal deviations, overuse of familiar FSs and so on (see Bishop, 2004; Meunier & Granger, 2008; Wray & Fitzpatrick, 2008). In the literature, FSs have been described as "the last and most challenging hurdle in attaining near nativelike fluency" (Spöttl & McCarthy, 2004, p. 191) because they are difficult even for advanced-level learners of English. For example, there is no grammatical explanation as to why "nice and easy" is idiomatic while "easy and nice" is not. The preference for the former is only a matter of an L1 speaker's habitual usage.

All over the world, researchers have been exploring ways of facilitating L2 learners' noticing of FSs. For example, Bishop (2004) and Szudarski and Carter (2016) have developed new ways of glossing FSs in written texts, and Granger (2011) has examined how classroom-based FS instruction may be improved (see also Meunier & Granger, 2008).

More recently, Lin (2012, 2016a, 2016b) proposes the need to maximize L2 learners' exposure to L2 spoken input as a way of facilitating their mastery of L2 FSs. It is argued that spoken input will be more conducive to FS acquisition than written input because FSs are prosodically salient. On paper, words are equally delineated by white spaces (Carter, 1987). There are no typographic cues to prompt learners to see that some words constitute relatively fixed chunks and, therefore, ought to be treated holistically (Bishop, 2004). However, the case is different when FSs are acquired through the auditory channel. Words that constitute relatively fixed chunks will tend to be articulated in an intonation unit (see Lin, 2010a, 2010b, 2013; Lin & Adolphs, 2009). This may prompt L2 learners to process and store these chunks holistically. Unfortunately, English as a Foreign Language (EFL) syllabi from many parts of the world tend to emphasize written input over spoken input (Lin, 2012). As a consequence, the imbalance between spoken and written input is severe. As Lin and Siyanova-Chanturia (2014) estimate, the proportion of spoken to written input in some EFL syllabi could be 1:9. On this basis, Lin (2014) advocates greater exposure to internet television outside the curriculum as a fundamental means of addressing this imbalance and maximizing EFL learners' exposure to spoken input in the L2. A new intelligent software package (Lin, 2016a, 2016b) has been developed to facilitate L2 FS acquisition through watching YouTube videos. The tool is designed to highlight pedagogically interesting FSs in L2 learners' self-chosen YouTube videos. It tracks each individual

learner's YouTube media history and then automatically generates cloze exercises based on users' performance statistics.

Despite the ongoing efforts to increase learners' exposure to spoken input in their L2 (Lin, 2016a, 2016b), whether this will be effective in promoting noticing and gains in knowledge of L2 FSs remains an open question. Although the aforementioned review demonstrates that prosody plays a vital role in many aspects of L1 acquisition, there is still some uncertainty about whether prosody is as significant for L2 acquisition. It is possible that the phonological representation of FSs is less prominent in the L2 than in the L1 mental lexicon. Even if prosody is as important to L2 acquisition as it is to L1, it cannot be assumed that an L2 learner's prosodic processing ability in the L2 will be as efficient as his/her prosodic processing ability in the L1. Therefore, he/she may not be able to fully exploit the prosodic cues in L2 spoken input and acquire L2 FSs in the same way that an L1 learner can. In fact, there is a body of literature which has revealed that learners' prosodic processing ability in their L2 typically fails to match the equivalent ability in their L1 (see the following section). If or when this is the case, the strategy of maximizing L2 learners' exposure to L2 spoken input in the hope that FSs will be noticed and acquired more readily through auditory input may only be effective if other measures are in place to enhance learners' L2 prosodic processing ability.

7. Learners' weakness in noticing and processing prosodic cues in second language input

In the literature on native listening (Cutler, 2012), a strand of research has investigated learners' L2 prosodic processing ability. These studies (Akker & Cutler, 2003; Broselow, Hurtig, & Ringen, 1987; Dupoux, Pallier, Sebastian, & Mehler, 1997; Pennington & Ellis, 2000) have shed light on the differences in the ways that native and non-native speakers handle prosodic cues in auditory input. It is now clear that non-native speakers are unable to match native speakers' efficiency in processing prosodic information, and their ability to handle L2 prosodic cues also tends to vary widely depending on their L1 background. In Akker and Cutler's (2003) study, for instance, Dutch learners and native speakers of English completed a phoneme-detection task designed to investigate the strategies they used for processing semantic focus and prosodic accent in the input. The results showed that Dutch learners of English were capable of exploiting the prosodic structure of the L2 spoken input and directing attention to accented words. However, the speed with which they were able to exploit prosodic cues and integrate them with other types of cues in the L2 input, such as semantic focus, was slower than their ability in their L1.

In studies involving learners whose L1 and L2 do not demonstrate similar prosodic structures, however, the conclusion is different. Pennington and Ellis (2000) tested the L2 prosodic processing of Cantonese learners of English using memory recognition tasks. In their experiments, subjects needed to decide whether they had heard identical sentences in a preceding exposure phase. The stimuli included foil items which differed from the originals in terms of either prosody or lexis. The results showed that, while these Cantonese learners of English had no problem with recognizing foil items that differed lexically from the originals, in the absence of explicit instruction, they were unable to recognize those that differed prosodically in terms of semantic focus (e.g., is HE driving the bus? vs. is he driving the bus), pragmatic interpretation (e.g., He's a good boy, isn't he? delivered in a falling vs. a rising tone), phrasing (e.g., The fight is over Fred vs. The fight is over, Fred), or internal structure (e.g., She's a lighthouse keeper vs. She's a light housekeeper). These learners' weakness in noticing and processing prosodic cues in the L2 input is thought to have arisen from the fact that their L1 and L2 (i.e., Cantonese and English) have fundamentally different prosodic structures. As Pennington and Ellis (2000) suggest, it is possible that details of the L2 prosodic system are not acquired until an advanced stage of language acquisition. In the case of L2 learners who have not achieved full competence in their L2 by adulthood, they transfer knowledge of their L1 prosodic patterns when processing L2 auditory input. The phenomenon of L1 transfer in prosodic processing is well-evidenced by studies involving speakers of tonal languages who are learning stress languages (Dupoux et al., 1997), as well as those involving speakers of stress languages learning tonal languages (Broselow et al., 1987). The extent to which this L1 transfer in prosodic processing affects the acquisition of L2 FSs from spoken input deserves further, empirical investigation.

Van Lancker (2003) conducted one of the first studies to compare native and non-native speakers' prosodic processing competence in relation to FSs. In the study, non-native speakers from mixed L1 backgrounds (e.g., Hebrew, Arabic, Spanish, Chinese, Korean, German, French, and Hungarian; 23 L1s in total) and native speakers of English listened to a number of prototypical English idioms embedded in sentences (e.g., She missed the boat; it broke the ice). These sentences appear semantically ambiguous on paper because they can be interpreted idiomatically or literally. However, the subjects were told that each sentence could be disambiguated prosodically. Their task was to make a forced-choice concerning whether each sentence had been read with an intended idiomatic or an intended literal meaning. Again, the results indicated that the prosodic processing competence of L2 learners was significantly worse than that of native speakers. Even highly fluent L2 learners had difficulties with perceiving the prosodic contrasts between the idiomatic and literal readings of the sentences.

With this in mind, there seems to be considerable empirical evidence for the wide gap between learners' L1 and L2 prosodic processing abilities. While prosodic processing ability probably begins prenatally for native speakers (see the biology studies cited earlier), for non-native speakers the extent of this ability depends on whether their L1 and L2 share similar prosodic structures. Where the prosodic structures of the L1 and L2 are similar, as in the case of the Dutch learners of English, L2 prosodic processing is possible, albeit with reduced efficiency. However, when the L1 and L2 prosodic structures differ considerably, as in the case of the Cantonese learners of English, even highly fluent L2 learners seem to struggle with extracting information about the semantic focus, the pragmatic interpretation, the phrasing and sentence structure from the prosodic cues.

8. Studies on the effect of spoken input on formulaic sequence acquisition in a second language

A decade ago, Fitzpatrick and Wray (2006) and Wray and Fitzpatrick (2008) conducted a series of interesting longitudinal studies to investigate how well L2 learners acquire FSs from spoken input. In these studies, intermediate to advanced-level Chinese and Japanese learners of English (whose L1s have prosodic structures different from that of English as the L2) were given the opportunity to identify practical, daily scenarios that were personally relevant. Authentic and formulaic dialogs appropriate to these scenarios were then prepared and audio-recorded onto CDs. The learners' task was to rote learn the formulaic responses to give in these dialogs in order to perform them in a role-play assessment. Despite having received spoken input, exposure to prosodic cues to formulaicity and repeated practice with the researchers, the learners were still unable to develop accurate memories of the formulaic responses and perform them as they appeared on the audio CDs. Instead, they introduced modifications of various sorts in their production (see Wray and Fitzpatrick, 2008, for details). The fact that these learners were unable to reproduce the formulaic responses intact means that they took an analytical (instead of the holistic, gestalt) approach when processing the spoken input. That the analytical mode was engaged is probably indicative of their lack of attention to prosodic cues which, as discussed earlier, should have facilitated the holistic processing of FSs. Since the learners' L2 prosodic processing ability was neither measured nor controlled, it was impossible to judge whether poor L2 prosodic processing ability had contributed to the poor learning outcome.

Whether poor L2 prosodic processing was the cause of the poor learning of FSs in these studies, other research does exist which suggests that learners' L2 prosodic processing can improve given instruction.

While Pennington and Ellis's (2000) first experiment showed that, unaided, even highly fluent Cantonese learners of English struggled with the extraction of information about semantic focus, pragmatic interpretation, phrasing, and sentence structure from prosodic cues, their second experiment showed that statistically significant improvements in L2 prosodic processing could occur if instruction was provided. In the phonological literature, there is a wealth of discussion about speech prosody instruction (Romero-Trillo, 2012; Trouvain & Gut, 2007), although most, if not all, of it has focused on improving the prosody of L2 learners' production, as opposed to ways of enhancing learners' processing of prosody during L2 perception. Considering both the evidence for non-native speakers' weaknesses in L2 prosody processing and its possible link with L2 FS acquisition, the demand for research of the latter type is substantial.

To shed light on the question of whether the imbalance between spoken and written input could be a reason for the lack in breadth of knowledge of L2 FSs that L2 learners seem to demonstrate, Lin (in preparation) is conducting an experiment to investigate: (1) whether spoken input is more conducive to L2 FS acquisition than written input and (2) whether spoken input with natural prosodic cues will be more conducive to L2 FS acquisition than spoken input with word-by-word prosody. At the same time, the study measures and controls for the effect of the subject's L2 processing ability, vocabulary size, and phonological short-term memory capacity. In the study, subjects were exposed to stimuli presented in one of three conditions: written input, spoken input with natural prosody, or spoken input with word-by-word prosody. To address the two research questions, the subjects' gains in knowledge of the FSs will be compared after controlling for the effect of L2 processing ability, vocabulary size, and phonological short-term memory capacity. The findings should provide an answer to key questions highlighted previously, including whether increasing L2 learners' exposure to spoken input is likely to facilitate L2 FS acquisition and whether prosody plays an important role in L2 FS acquisition.

9. Conclusions and future directions

To conclude, this chapter has offered a critical review of the role of prosody in the acquisition of L1 and L2 FSs. Considering the fact that 90% of an ordinary person's daily linguistic encounter in his/her L1 is spoken (Carter, personal communication, 2012) and that L1 FSs are acquired and accessed through the auditory channel during extensive exposure to spoken communication, this chapter proposes that FSs be reconceptualized (and redefined) as strings of sounds rather than strings of words. If this

reconceptualization is to take place, it is important that a deeper understanding is gained of the phonological representation of FSs in the mental lexicon.

In fact, empirical evidence from biology and child language research already points to the fundamental role of prosody in FS acquisition in the case of child L1 learners. Foetuses are biologically predisposed to rely on prosodic cues to acquire their L1. Natural speech also contains prosodic patterns that guide the segmentation of the input and influence the unit of lexical storage in the brain. Thus, from a purely statistical point of view, these prosodic patterns could prompt the acquisition of L1 FSs in the same way that they prompt the acquisition of L1 grammar.

Compared with L1 FS acquisition, L2 FS acquisition appears to be much more complex. Studies have investigated why the acquisition of FSs seems surprisingly difficult for L2 learners. One of the proposals traces the cause of the problem to L2 learners' lack of exposure to L2 spoken input and suggests that increased exposure to prosodic cues should prompt learners to notice and acquire FSs. This proposal also predicts that maximizing L2 learners' exposure to L2 spoken input outside the curriculum may help to facilitate L2 FS acquisition. In this chapter, the proposal was critically examined in the light of recent findings from studies on L2 FS acquisition and native listening, concluding that perhaps maximizing L2 spoken input will facilitate L2 FS acquisition only on the condition that L2 learners have already achieved a certain level of L2 prosodic processing competence. This suggestion is currently being tested in a control experiment.

In terms of future directions, it is noteworthy that previous studies have often considered FSs from a lexical perspective only and overlooked the fact that prosody is vital to the acquisition, processing, and use of spoken FSs. While corpus linguists continue to examine the prosodic patterns of FSs and how they affect contextual meanings, the role of prosody in L1 and L2 FS acquisition remains an uncharted territory. Many fundamental questions are still awaiting empirical testing. For example, while it is clear that spoken input dominates in L1 acquisition and written input dominates in L2 acquisition, it is unclear whether this difference in the dominant mode of input shapes the ways that L1 and L2 FSs are represented in the mental lexicon. If the mode of input does appear to be a significant factor of success in L2 FS acquisition, then it would be useful to note the extent to which that success is moderated by the learner's auditory processing ability. In relation to this, there is the question of whether learners with better L2 prosodic processing ability will master FSs faster than learners with poor L2 prosodic processing ability. The effect of speech prosody

instruction on learners' success in L2 FSs acquisition is also an important area for study. If these questions can be answered in the future, we will be one step closer to tackling L2 learners' problems with acquiring L2 FSs.

Acknowledgement

The preparation for this paper was supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project number: PolyU 25612116) and a grant from the Hong Kong Polytechnic University (Project number: 1-ZVET).

References

Abdelli-Baruh, N., Yang, S.-y., Ahn, J. S., & van Lancker, D. (2007). *Acoustic cues differentiating idiomatic from literal expressions across languages*. Paper presented at the American Speech and Hearing Association Convention, Boston, November 15–17, 2007.

Aijmer, K. (1996). *Conversational routines in English*. London: Longman.

Akker, E., & Cutler, A. (2003). Prosodic cues to semantic structure in native and nonnative listening. *Bilingualism: Language and Cognition, 6*(2), 81–96.

Ashby, M. (2006). Prosody and idioms in English. *Journal of Pragmatics, 38*(10), 1580-1597.

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.

Bishop, H. (2004). The effect of typographic salience on the look up and comprehension of unknown formulaic sequences. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 227–248). Amsterdam: John Benjamins.

Bloom, L. (1973). *One word at a time: The use of single word utterances before syntax*. The Hague: Mouton.

Bolinger, D. (1965). On certain functions of accents A and B. In I. Abe & T. Kanekiyo (Eds.), *Forms of English* (pp. 57–66). Cambridge, MA: Harvard University Press.

Branigan, G. (1979). Some reasons why successive single word utterances are not. *Journal of Child Language, 6*, 411–421.

Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology* (pp. 350–361). Cambridge, MA: Newbury House.

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

Carter, R. (1987). *Vocabulary: Applied linguistic perspectives*. London: Allen and Unwin.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.

Cutler, A. (2012). *Native listening language experience and the recognition of spoken words*. Cambridge, MA: MIT Press.

DeCasper, A. J., & Spence, M. J. (1986). Prenatal maternal speech influences newborns' perception of speech sounds. *Infant Behavior and Development, 9*(2), 133–150.

Dechert, H. W. (1983). How a story is done in a second language. In C. Faerch & G. Kasper (Eds.), *Strategies in interlanguage communication* (pp. 175–195). London: Longman.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A Destressing "Deafness" in French? *Journal of Memory and Language, 36*(3), 406–421.

Ferrari, G. A., Nicolini, Y., Demuru, E., Tosato, C., Hussain, M., Scesa, E., . . . Ferrari, P. F. (2016). Ultrasonographic investigation of human fetus responses to maternal communicative and non-communicative stimuli. *Frontiers in Psychology, 7*, 354–363.

Fisher, C., & Tokura, H. (1996). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 343–363). Hillsdale, NJ and England: Lawrence Erlbaum.

Fitzpatrick, T., & Wray, A. (2006). Breaking up is not so hard to do: Individual differences in L2 memorization. *The Canadian Modern Language Review, 63*(1), 35–57.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology, 15*(4), 411–458.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language and speech, 15*(2), 103–113.

Granger, S. (2011). From phraseology to pedagogy: Challenges and prospects. In T. Herbst, S. Faulhaber, & P. Uhrig (Eds.), *The phraseological view of language* (pp. 123–146). Berlin: Mouton de Gruyter.

Iwasaki, S., & Tao, H. (1993). *A comparative study of the structure of the intonation unit in English, Japanese, and Mandarin Chinese*. Paper presented at The 67th Annual Meeting of the Linguistic Society of America, Los Angeles, California, January 7–10, 1993.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D. G., Kennedy, L. J., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology, 24*, 252–293.

Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., & Wright Cassidy, K. (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language, 16*, 55–68.

Knowles, G. (1991). Prosodic labelling: The problem of tone group boundaries. In S. Johansson & A.-B. Stenström (Eds.), *English computer corpora* (pp. 149–163). Berlin: Mouton de Gruyter.

Lin, P. (2010a). The phonology of formulaic sequences: A review. In D. Wood (Ed.), *Perspectives on formulaic language* (pp. 174–193). London: Continuum.

Lin, P. (2010b). *The prosody of formulaic language* (Unpublished doctoral dissertation). University of Nottingham, Nottingham, UK.

Lin, P. (2012). Sound evidence: The missing piece of the jigsaw in formulaic language research. *Applied Linguistics, 33*(3), 342–347.

Lin, P. (2013). The prosody of idiomatic expressions in the IBM/Lancaster Spoken English Corpus. *International Journal of Corpus Linguistics, 18*(4), 561–588.

Lin, P. (2014). Investigating the validity of internet television as a resource for acquiring L2 formulaic sequences. *System, 42*(1), 164–176.

Lin, P. (2015). *It's not what you say; It is the way that you say it: The need to specify the intonation of idioms and phraseology in learner dictionaries*. Paper presented at The Asian Association for Lexicography (ASIALEX) conference, The Hong Kong Polytechnic University, Hong Kong, June 25–27, 2015.

Lin, P. (2016a). *How many exposures do learners need to learn a new English phrase*. Paper presented at the Doing Linguistics with Big Data: The 20th Workshop on Linguistics and Language Processing, Kyung Hee University, Seoul, May 29, 2016.

Lin, P. (2016b). *Second language vocabulary acquisition in the age of Internet Television and Social Media*. Paper presented at the Department of English Seminar, The Hong Kong Polytechnic University, November 7, 2016.

Lin, P. (2018). *The prosody of formulaic sequences: A corpus and discourse approach*. London: Continuum.

Lin, P. (in preparation). Investigating the optimal mode of input for the acquisition of second language formulaic sequences.

Lin, P., & Adolphs, S. (2009). Sound evidence: Phraseological units in spoken corpora. In A. Barfield & H. Gyllstad (Eds.), *Researching collocations in another language* (pp. 34–48). Basingstoke: Palgrave Macmillan.

Lin, P., & Siyanova-Chanturia, A. (2014). Internet television for L2 vocabulary learning. In D. Nunan & J. C. Richards (Eds.), *Language learning beyond the classroom* (pp. 149–158). London: Routledge.

Mampe, B., Friederici, A. D., Christophe, A., & Wermke, K. (2009). Newborns' cry melody is shaped by their native language. *Current Biology, 19*, 1–4.

Meunier, F., & Granger, S. (Eds.). (2008). *Phraseology in foreign language learning and teaching*. Amsterdam: John Benjamins.

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

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal, 84*(3), 372–389.

Peters, A. M. (1974). *The beginnings of speech*. Papers and Reports on Child Language Development, Stanford University, 8, 26-32.

Peters, A. M. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language, 53*(3), 560–573.

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

Raupach, M. (1984). Formulae in second language speech production. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 114–137). Tubingen: Gunter Narr.

Romero-Trillo, J. (Ed.) (2012). *Pragmatics and prosody in English language teaching*. Dordrecht: Springer.

Schuetze-Coburn, S. (1994). *Prosody, syntax, and discourse pragmatics: Assessing information flow in German conversation* (Unpublished doctoral dissertation). University of California, Los Angeles.

Scott, D. R. (1982). Duration as a cue to the perception of a phrase boundary. *The Journal of the Acoustical Society of America, 71*(4), 996–1007.

Siyanova-Chanturia, A., & Lin, P. (2017). Production of ambiguous idioms in English: A reading aloud study. *International Journal of Applied Linguistics*, Early View Article, 1–13.

Spöttl, C., & McCarthy, M. (2004). Comparing knowledge of formulaic sequences across L1, L2, L3, and L4. In N. Schmitt (Ed.), *Formulaic sequences* (pp. 191–226). Amsterdam and Philadelphia: John Benjamins.

Szudarski, P., & Carter, R. (2016). The role of input flood and input enhancement in EFL learners' acquisition of collocations. *International Journal of Applied Linguistics, 26*(2), 245–265.

Trouvain, J., & Gut, U. (Eds.). (2007). *Non-native prosody: Phonetic description and teaching practice*. Berlin: Mouton de Gruyter.

van Lancker, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics, 24*(1), 45–57.

van Lancker, D., & Canter, G. J. (1981). Idiomatic versus literal interpretations of ditropically ambiguous sentences. *Journal of Speech and Hearing Research, 46*(1), 64–69.

van Lancker, D., Canter, G. J., & Terbeek, D. (1981). Disambiguation of ditropic sentences acoustic and phonetic cues. *Journal of Speech and Hearing Research, 24*(3), 330–335.

Wermke, K., & Mende, W. (2009). Musical elements in human infants' cries: In the beginning is the melody. In O. Vitouch & O. Ladinig (Eds.), *Musicae scientiae, special issue on music and evolution* (pp. 151–173). Brussels: Presses Universitaires de Bruxelles.

Wichmann, A. (2000). *Intonation in text and discourse*. Harlow: Longman.

Wood, D. (2012). *Formulaic language and second language speech fluency*. London: Continuum.

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

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

Wray, A., & Fitzpatrick, T. (2008). Why can't you just leave it alone? Deviations from memorized language as a gauge of nativelike competence. In F. Meunier & S. Granger (Eds.), *Phraseology in foreign language learning and teaching* (pp. 123–147). Amsterdam: John Benjamins.

Yang, S.-y., Ahn, J. S., & Van Lancker, D. (2010). The perception and acoustic features of Korean ditropic sentences. *Journal of the Acoustical Society of America, 127*(3), 1955–1955.