

Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2*

GARETH CARROL
KATHY CONKLIN
University of Nottingham

(Received: March 2, 2013; final revision received: November 6, 2013; accepted: November 19, 2013; first published online 4 February 2014)

Monolingual speakers show priming for idiomatic sequences (e.g. a pain in the neck) relative to matched controls (e.g. a pain in the foot); single word translation equivalents show cross-language activation (e.g. dog–chien) for bilinguals. If the lexicon is heteromorphic (Wray, 2002), larger units may show cross-language priming in the same way as single words. We used the initial words of English idioms (e.g. to spill the . . . beans) and transliterated Chinese idioms (e.g. draw a snake and add . . . feet) as primes for the final words in a lexical decision task with high proficiency Chinese–English bilinguals and English monolinguals. Bilinguals responded to targets significantly faster when they completed a Chinese idiom (e.g. feet) than when they were presented with a matched control word (e.g. hair). The results are discussed in terms of conceptual activation and lexical translation processes, and are also incorporated into a dual route model of formulaic and novel language processing.

Keywords: bilingualism, dual route processing, fast automatic translation, formulaic language, idioms

Introduction

Formulaic language (idioms, speech formulae, clichés, etc.) is no *flash in the pan*. The definition of formulaic language used here is taken from Wray and Perkins (2000): “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”. Such sequences account for between a third and a half of spontaneous discourse (Erman & Warren, 2000; Foster, 2001). They contribute to speaker fluency (Pawley & Syder, 1983), facilitate real-time communication (Code, 1994) and reduce demands on working memory (Conklin & Schmitt, 2008). They present a particular challenge to non-native speakers, as they are both an important part of native-like competence and one of the hardest aspects of a language to master. Cieślicka (2006) suggested that a better understanding of how non-native speakers acquire and use formulaic language should be a key goal of modern psycholinguistic and applied linguistic research.

A dual route model (Van Lancker Sidtis, 2012; Wray, 2002; Wray & Perkins, 2000) can provide a means of describing formulaic language processing in native speakers. In this view, two approaches to processing are available to speakers: frequent, familiar phrases are stored

in long-term memory and can be accessed or retrieved directly, while novel phrases are computed using a words-and-rules approach. It is important to note that whilst the advantage for formulaic language is often referred to as ‘retrieval’ or ‘holistic storage’, we use such terms only as a convenient shorthand to describe attested processing differences between formulaic and novel language. The processing advantage for formulaic language could reflect the unitary storage of whole forms, but equally it could arise from the simultaneous activation of the component parts of a phrase or the priming of multiple combinations via the base components (Wray, 2012, p. 234). Throughout this paper, RETRIEVAL refers to access to the components and meaning of a familiar phrase in a way that is quicker than computing a comparable control phrase. This offers formulaic sequences an advantage over matched novel language as it is a qualitatively different and fundamentally faster process than computation (Tabossi, Fanari & Wolf, 2009).

Whilst a processing advantage is clear for native speakers (see Van Lancker Sidtis, 2012 or Wray, 2012 for reviews), formulaic language processing in non-native speakers remains comparatively unexplored, particularly in terms of how the bilingual lexicon might accommodate two distinct processing routes when more than one language is involved. We address this question by investigating how sequences that would be formulaic in a first language (L1) are processed when they are encountered in a second language (L2). For example, if a French–English bilingual speaker encounters the English sequence *howl with the wolves*, will he or she recognise and retrieve the underlying French idiom

* This research was supported by a Sir Francis Hill Postgraduate Scholarship awarded to the first author. We are grateful for the comments of three anonymous reviewers on an earlier draft of this article, and to Ping Li for his comments and suggestions.

Address for correspondence:

Gareth Carrol, School of English, The University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom
gareth.carrol@nottingham.ac.uk

hurler avec les loups (comparable to the English idiom *follow the crowd*)? If formulaic language represents the storage and association of frequently encountered forms, then we might expect such units to be language-specific: encountering a known sequence in an unfamiliar (L2) form should show no advantage over a matched control phrase. If an advantage is observed for unfamiliar translated forms, this would imply some level of L1–L2 interaction in the processing of formulaic sequences: since the configuration *howl with the wolves* does not exist in English, any processing advantage cannot be located at a purely lexical level in the L2. Despite the wealth of research into formulaic language to date, no study has investigated this question.

Evidence for a dual route model

Formulaic language is processed more quickly than matched novel language by native speakers. This has been consistently demonstrated for idioms (Gibbs, 1980; Swinney & Cutler, 1979), collocations (Durrant, 2008), corpus-derived multiword units (Ellis, Simpson-Vlach & Maynard, 2008; Jiang & Nekrasova, 2007) and multiword lexical verbs (Isobe, 2011). This formulaic/novel discrepancy is supported by widespread evidence of different patterns in the brain's electrophysiological response (ERP) to such stimuli (Siyanova, 2010; Tremblay & Baayen, 2010; Vespignani, Canal, Molinaro, Fonda & Cacciari, 2009), and by evidence of different patterns of performance for left-hemisphere and right-hemisphere brain-damaged patients (Code, 2005; Van Lancker Sidtis & Postman, 2006). There is a wealth of psychological and neurological evidence to support two distinct routes for language processing according to the nature of the material being processed (Van Lancker Sidtis, 2012). The retrieval route can be used for previously encountered phrases; recognition of such a phrase will provide access to the underlying canonical form, its conventional meaning and its pragmatic conditions of usage. Subjective familiarity ultimately determines whether the direct route is available, and the dual route model can be seen as a race rather than an either/or choice: computation still takes place for known phrases, but direct access returns the same results more quickly (and in the case of figurative language is more likely to return the intended meaning than literal analysis), whereas for unfamiliar phrases only the computation route is available. Tabossi et al. (2009) showed that familiarity was the main driver of the processing advantage for both non-compositional formulaic sequences (idioms) and compositional units (clichés).

The present experiment uses idioms, which are “evidently formulaic” (Wray, 2008, p. 28). Non-decomposable idioms, or what Grant and Bauer (2004, p. 40) call “core idioms”, present a particular problem

because they are, at a surface level, incomprehensible, opaque and gnomic. Crucially, idioms are ubiquitous in discourse and their figurative meanings are processed without difficulty by native speakers. For the present purposes, it is their clear formulaicity that is important; we take this to mean that idioms have an attested L1 citation form that will be recognised and understood by a majority of native speakers, and the question is whether the advantage offered by direct access is based primarily on recognition of form. Given the importance of familiarity, it seems logical that presentation in a non-native language should impair recognition of the formulaic sequence. However, idioms are often much more flexible than people assume (Schmitt, 2005), and native speakers generally have little trouble dealing with non-standard and creative idioms provided they are not too far removed from the citation form (Omazic, 2008). Hence while early models (e.g. Bobrow & Bell, 1973; Gibbs, Nayak & Cutting, 1989; Swinney & Cutler, 1979) broadly described idioms as single entries in the lexicon, more recent hybrid accounts (e.g. Cacciari & Glucksberg, 1991; Cacciari & Tabossi, 1988; Cutting & Bock, 1997; Sprenger, Levelt & Kempen, 2006) have attempted to incorporate the syntactic and lexical flexibility of idioms, as well as attempting to explain the finding that both the literal meanings of individual words and the idiomatic meaning of the whole phrase seem to be available during idiom processing. Idioms may therefore be simultaneously compositional and non-compositional (Kuiper, van Egmond, Kempen & Sprenger, 2007), which argues against a view that they are represented as single, unanalysable units. Instead, they may represent configurations with distributed meanings in the lexicon, according to the Configuration Hypothesis proposed by Cacciari and Tabossi (1988), or they may represent separate lexical-conceptual entries – what Sprenger et al. (2006) call *SUPERLEMMAS* – that are accessible via the component words. A dual route model therefore allows idioms to be directly accessed, which unlocks both their lexical components and the phrasal figurative meaning. Figure 1 shows a representation of a dual route model for the English idiom *flog a dead horse* (meaning “to persevere pointlessly with a task that will have no positive outcome”).

When *flog a* is encountered, obligatory analysis and computation begins, as in (1), until the recognition point is reached, which is what Cacciari and Tabossi (1988, p. 678) refer to as the “key” of the idiom. Logically this must be the default approach, because it is only by encountering enough of the component parts of a known phrase that it can be recognised and unlocked. For any sequence, therefore, the computation route is available, but previously encountered phrases, once the recognition point has been reached, may also be accessed directly. Hence encountering the combination of *flog, a*

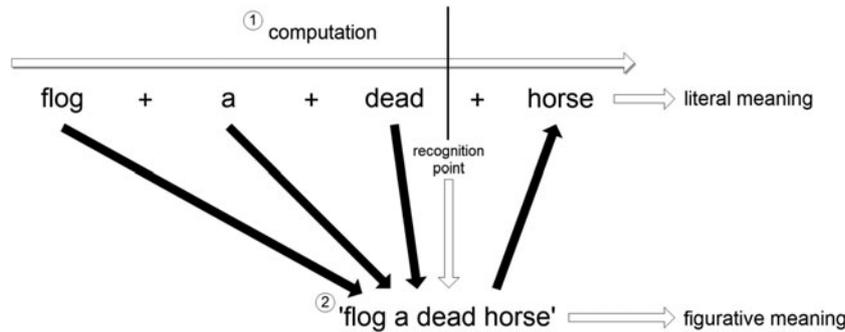


Figure 1. Dual route model for the English idiom *flog a dead horse*. The two routes represented are obligatory analysis and computation according to the individual words and grammar (1) and direct recognition and activation of the lexical-conceptual configuration of the idiom (2). Black arrows represent associative links between components and white arrows represent processes.

and *dead* triggers enough associations to activate the known configuration of the idiom. As *horse* is a part of this configuration it is automatically activated before it is encountered as part of the computation. The retrieval route is therefore faster because, in the case of idioms with an early recognition point, the final components are activated before they are encountered via compositional analysis. Because *horse* has been activated as part of the full idiom, if this is the next word to be encountered then it will be processed more quickly, but if another word appears (e.g. in a control phrase like *flog a dead beast*), processing will continue compositionally.

For idioms with late recognition points (i.e. only after the final word, as in *kick the bucket*), the temporal advantage is perhaps not as clear. However, such idioms might still have a processing advantage, for two reasons. Firstly, encountering *kick the* should activate *bucket* to some extent, even though the idiom has not been fully recognised, especially if the context is supportive of the idiomatic usage. While unequivocal recognition might not occur until the final word has been seen, the idiom is likely to be already activated at least to some degree. This is congruent with Sprenger et al. (2006), who suggested that idiom recognition is contingent on reaching a threshold of activation based on encountering progressively more components of a phrase. This threshold may therefore represent confirmation of the idiom, but each component will contribute something toward idiom activation. Secondly, once the final word of an idiom has been encountered it will be activated both as part of the idiom and as part of a computational analysis. Hence *bucket* would be activated by both routes simultaneously, providing an advantage relative to a control phrase (e.g. *kick the packet*), which would only be activated via the computation route.

For any novel phrase, only the computational route is available. Until an idiom (or other formulaic sequence) has been encountered with enough frequency to form associative links between components and therefore create

configurations, no direct access will be available, so non-native speakers are unlikely to be able to use the direct access route until a certain level of proficiency has been reached. There is evidence that once they have encountered formulaic sequences in the L2 with enough regularity, non-native speakers demonstrate the same advantage as native speakers (Isobe, 2011; Jiang & Nekrasova, 2007). There is therefore no fundamental difference in how native and non-native speakers process formulaic language, but there is likely to be a large discrepancy in the strength of associations available to trigger direct access. This means that non-native speakers are more likely to process L2 formulaic language compositionally and to encounter problems when this does not produce intelligible results (e.g. in the case of entirely opaque idioms).

Translated idioms and cross-language priming in bilinguals

The dual route model is less clear in its predictions for translated idioms. There is widespread evidence to support priming effects in bilinguals for single words (Chen & Ng, 1989; de Groot & Nas, 1991). Translation equivalents, in particular (e.g. *dog/chien*), consistently show cross-language facilitation for bilingual speakers, which Wang (2007) suggests is a reflection of their shared conceptual representations. Therefore, there is clearly some level of interaction between single word representations in different languages. However, an important consideration is that such associative links are likely to be highly asymmetrical. Whilst a French–English bilingual is likely to have connected representations for the L1 and L2 forms of *hurler–howl*, *avec–with* and *les loups–the wolves*, the lexical associations between these items that unlock the underlying idiom should exist only in the L1 (French). Use of the direct route across languages may therefore require mediation via a conceptual level, whereby the individual L2 forms activate their conceptual

representations and the associations at this level trigger the concept underlying the idiom.

Studies have shown cross-language effects at a level above the single word, which would lend support to a conceptual basis for the dual route. For example, Japanese–English bilinguals responded more quickly to unconnected English word pairs that were translations of L1 Japanese collocations (e.g. *forgive marriage*) than to unrelated control pairs (Ueno, 2009). Wolter and Gyllstad (2011) found similar results for Swedish–English bilinguals, with facilitation for English word pairs that formed congruent collocations in English and Swedish (e.g. *give an answer*, which is the word-for-word translation equivalent of *ge ett svar*) relative to English-only collocations (e.g. *pay a visit*, where the Swedish translation equivalent for *pay* cannot be used idiomatically in a phrase like **betala ett besök*). Both studies concluded that language non-selective conceptual associations can drive lexical effects in the L2. Given the evidence for cross-language effects in single words and collocations, it seems logical that larger units (idioms) may demonstrate similar effects. The current experiment will explore that question by investigating whether Chinese–English bilinguals, relative to matched controls, show any facilitation for Chinese idioms that have been translated into English.

Chinese idioms

Chinese has a large set of homogenous idioms that are ideal for the purposes of the current investigation. *Chengyu* (“fixed expressions”) generally consist of four fixed characters, allowing no semantic substitution or syntactic flexibility without destroying the integrity of the idiom. Around 97% of all *chengyu* conform to the four-character structure (Liu, Li, Shu, Zhang & Chen, 2010). They are generally semantically opaque, and many refer to a folk story or historical event. Understanding the intended meaning is therefore contingent on either knowing the underlying story or learning the arbitrary idiomatic meaning of the sequence.

Chengyu are formulaic units in Chinese (Simon, Zhang, Zang & Peng, 1989; Zhou, Zhou & Chen, 2004) and have been shown to hold the same processing advantage as English idioms. This has been demonstrated through shorter reaction times to *chengyu* than to matched control sequences (Liu et al., 2010; Zhang, Yang, Gu & Ji, 2013) and through ERP data showing different responses for idiomatic and matched non-idiomatic sequences (Zhou et al., 2004; Liu et al., 2010). Chung, Code and Ball (2004) described similar patterns of impairment in individuals with aphasia for Chinese and English speakers, i.e. differential performance in formulaic vs. novel language. This evidence supports a dual route mechanism for language processing in Chinese, just as in English.

For the current investigation, sets of English and Chinese idioms were prepared to explore the responses of Chinese–English bilinguals to formulaic language from the L2 (English idioms) and translated from the L1 (translated Chinese idioms). The responses of English native speakers were also collected for comparison. A lexical decision task was used to compare responses to idioms and matched controls for both languages. If the processing advantage for idioms is based on recognition and retrieval of known forms, we would expect no advantage for translated idioms for the non-native speakers. We would also expect any advantage for English idioms to be driven by proficiency. Native speakers should show an advantage for L1 (English) idioms vs. controls and no difference for translated idioms and controls.

Methodology

Participants

Nineteen native speakers of English (with no experience of learning Mandarin) and 19 non-native speakers of English took part in the experiment for course credit. The non-native speakers all had Mandarin Chinese as their first language and were students undertaking a year of study abroad at an English university. A summary of the non-native participants is shown in Table 1. All non-native participants were asked to complete a short language background questionnaire and a vocabulary test (modified from Nation & Beglar, 2007). The test presented a series of vocabulary items, each embedded in a short, context-neutral sentence (e.g. “Poor: we are poor”) and participants were asked to choose from five possible definitions: a correct response, three distractors and a “don’t know” option. The test included two items each from the first ten British National Corpus (BNC) word lists (the 10,000 most frequent word families in English) to give a total proficiency score out of 20. This was augmented with any potentially unknown vocabulary items that appeared in the online experiment (e.g. in the Chinese idiom “a horse does not stop its hooves”, “hooves” might be an unfamiliar English word so was included in the test to verify whether it was known to the participants). Any words that appeared in the stimulus phrases (primes or targets) that were outside the 2,000 most frequent word families in English were included in the test. If any participant failed to choose the correct response for a word from one of these idioms, the idiom containing that word was removed from the analysis for that participant. This meant that 33 words were included in the modified vocabulary test, to give a total of 53 items. The language background questionnaire asked participants to provide information about the length of time they had been studying English and to estimate their English proficiency in reading, writing, listening and

Table 1. Summary of non-native speakers' age, years of studying English, self-rating of English proficiency, estimate of usage and vocabulary test scores.

	Age	Years studying English	Reading	Writing	Speaking	Listening	Usage	Vocab
Mean	20.8	10.2	2.9	2.9	2.8	2.8	34.6	11.7
Range	19–22	5–15	2–4	2–4	1–5	2–4	26–46	7–16

Note: Reading, Writing, Speaking and Listening self-rated out of five (1 = Poor, 2 = Basic, 3 = Good, 4 = Very good, 5 = Excellent); Usage is an aggregated estimate of how frequently participants use English in their everyday lives (score out of 50 based on ten measures such as reading for pleasure or watching TV); Vocab is a modified Vocabulary Levels Test with a total score out of 20.

speaking (score out of five for each discipline). They were also asked to indicate how often they used English in their everyday lives (speaking to friends, attending lectures, reading in English for pleasure, watching TV, etc.). Each of these was scored on a five-point Likert scale and then aggregated into an overall usage rating (ten measures, each scored out of five to give an overall score out of 50). Both the vocabulary test and language background questionnaire were administered after the online experiment to eliminate any danger of repetition effects.

Materials

The stimulus materials consisted of English idioms, English control phrases, translated Chinese idioms and translated Chinese control phrases. Control items were formed by replacing the final word of the corresponding idiom with an unrelated but logical alternative (e.g. *spill the beans* vs. *spill the chips*).

English idioms were selected from the *Oxford Learner's Dictionary of English Idioms* (Warren, 1994) from those with a monosyllabic final word that was either a noun (e.g. *jump the gun*) or, in one case, an adjective (*the coast is clear*). As recognition of familiar phrases was the main concern, no distinction was made between types of idioms, for example in terms of the core idioms, figuratives and ONCEs (one non-compositional element) classification developed by Grant and Bauer (2004). To ensure that the stimuli were generally well known, all English idioms were normed on a population of native speakers using a cloze test (i.e. to reveal a secret is to *spill the...*) and were correctly completed by at least 90% of respondents. Mean length of the final word of each idiom (the target) was 4.5 letters and mean occurrence in the BNC was 21 (per 100 million words). Control items were created by selecting an alternative final word that was matched with the original for part of speech, length and frequency. Independent samples *t*-tests showed no difference between the idioms and the control items for length ($p = .69$) or frequency ($p = .43$). All alternative phrases showed a phrase frequency of 0 in the BNC.

Chinese idioms were initially selected from the *Dictionary of 1000 Chinese idioms* (Lin & Leonard,

2012). Only idioms where a literal translation provided a plausible English sequence with identical word order were considered, e.g. 畫蛇添足 – “draw-snake-add-feet” = “draw a snake and add feet”, meaning “to ruin something by adding over-elaborate and unnecessary detail”. The final character had to have a monosyllabic single word translation equivalent in English. The 20 that most closely matched the English idioms in length and frequency of the final word were retained. Four Chinese speakers confirmed that all were well known (all recognised by 4/4 speakers); this was not used as a strict norming test as all idioms were later assessed for subjective familiarity following the online experiment, but was intended simply to ensure that the idioms were likely to be recognised by the majority of participants. Translations were initially taken from the gloss provided by the *Dictionary of 1000 Chinese idioms*. Because the intention was to recreate the form of each idiom as closely as possible, the translations were checked character by character using two different online translation engines (Google Translate and Online Chinese Tools). In this way it was possible to get good agreement on the best literal translation of each character. The translations were finally verified by three native speakers of Chinese, who agreed that they were accurate representations of the Chinese originals. The mean length of the final word of each translated idiom was 4.7 letters and all translated Chinese idioms showed a phrase frequency of 0 in the BNC. Control items were created by replacing the final word of each translated idiom with a word matched for part of speech, length and frequency that formed a plausible sequence (e.g. *draw a snake and add hair*). Independent samples *t*-tests showed no difference between the idioms and the control items for length ($p = .73$), and a marginal difference for raw frequency ($p = .09$), although there was no difference for the frequency band of the items ($p = .77$). All alternative phrases showed a phrase frequency of 0 in the BNC.

A set of literal English phrases was constructed to act as filler material. All were literally plausible, grammatical English phrases (e.g. *carry the tray*) and each showed a phrase frequency of 0 in the BNC. Targets were monosyllabic and matched the idiom conditions for length (mean = 4.5 letters) and frequency. Non-word targets were created to make an equal number of word/non-word

responses. All non-words were taken from the ARC non-word database (Rastle, Harrington & Coltheart, 2002), conformed to the phonotactic rules of English and were matched with the other conditions for length (mean = 5.0 letters). Primes for the non-words were a mix of unused items from the English idiom, Chinese idiom and English literal conditions.

All idioms were assessed for compositionality using a method adapted from Tabossi, Fanari and Wolf (2008). English native speakers ($n = 16$) were presented with the English and Chinese idioms and a literal paraphrase of each (e.g. *to spill the beans* means “to reveal a secret”). Participants were asked to judge on a seven-point Likert scale how easily they thought the meaning of the idiom could be mapped onto the literal paraphrase. The mean rating for English idioms was 4.6/7, S.D. = 0.91 and for Chinese idioms was 3.8/7, S.D. = 1.55. In addition, the Chinese idioms were presented in the original Chinese characters to a group of 12 Chinese native speakers (who did not take part in the online task), who were asked to judge on a seven-point Likert scale how much they thought the individual characters contributed to the idiomatic meaning. The mean rating by Chinese native speakers was 5.5/7, S.D. = 1.10. There was no correlation between the two sets of compositionality judgements ($r = .33$; $p = .16$), and the discrepancy is itself a point of interest. In some ways the English speakers’ ratings may represent a “purer” measure of compositionality for the Chinese idioms, as they have no knowledge of the folk story or historical event that underpins the idiomatic meaning; their judgements are therefore based entirely on how clearly the linguistic information contributes to the figurative meaning of the Chinese idiom. In contrast, the Chinese native speakers may see the idioms as more transparent as a result of knowing the underlying stories. Our analysis will include both variables to see if either measure has an effect on response times (RTs).

The stimuli were divided into two counterbalanced lists with an idiom and its control appearing on opposite lists. Each participant saw ten English idioms, ten English controls, ten translated Chinese idioms, ten Chinese controls, 20 English filler items and 60 items with non-word targets (see Table 2). Independent samples t -tests showed no significant differences between the lists in target length ($A = 4.55$; $B = 4.55$; $p = 1$), target frequency ($A = 9860$; $B = 10101$; $p = .95$) or phrase frequency (English idioms only: $A = 20.8$; $B = 21.8$; $p = .86$). Care was also taken to ensure that the idioms on each list were balanced for compositionality, including both the scores by English native speakers (for both sets of idioms) and Chinese native speakers (for Chinese idioms only). The lists showed no significant differences for native speaker ratings of English idioms ($A = 4.5$; $B = 4.7$; $p = .52$), English native speaker ratings of translated Chinese idioms ($A = 3.3$; $B = 4.3$; $p =$

Table 2. Example of stimulus materials for each condition.

Condition	Prime	Target
English idiom	On the edge of your	seat
English control	On the edge of your	plate
Chinese idiom	Draw a snake and add	feet
Chinese control	Draw a snake and add	hair
Control phrase + real word	Put it in your	dish
Control phrase + non-word	Cut a long story	tealth

.17) or Chinese native speaker ratings of Chinese idioms ($A = 5.3$; $B = 5.7$; $p = .43$). Stimulus materials from the experimental conditions are available as Supplementary Online Materials.

Procedure

The experiment was conducted in a quiet laboratory using E-Prime (v.1.4.1.1) to present participants with the prime phrases and the target words for the lexical decision task. Reading of the prime phrases was self-paced: participants were asked to read the phrase as quickly as possible, then to press a button to advance once they had finished reading. A self-paced protocol was adopted to allow for the variation in reading time between native and non-native speakers. Once the prime disappeared a line of asterisks appeared on screen. After 250 ms this disappeared and the target was presented. Participants used a serial response button box to indicate whether the target was a real English word (YES/NO). Accuracy and RTs were recorded. The task was explained to each participant via on-screen instructions and two examples and six practice items were presented. The stimuli were then presented in random order until each participant had seen all 120 items.

Following this, participants were asked to rate all idioms for how familiar they considered them to be. For native speakers all idioms were presented in English. Participants used a seven-point Likert scale to indicate familiarity with each phrase. For non-native speakers the English idioms were presented in English and the Chinese idioms were presented in the original Chinese characters. Participants were again asked to rate how familiar they were with each phrase on a seven-point Likert scale.

Results and analysis

Two non-native speakers were removed from the analysis: both had a large number of extreme RTs, suggesting that either they were not engaging in the task or that the English task was too difficult for them. This left data

Table 3. Mean response times in msec and standard deviations (in brackets) for native and non-native speakers in each of the four experimental conditions. Non-native speaker values are vocabulary-adjusted (any unknown items removed).

	English idioms	English controls	Chinese idioms	Chinese controls
Native speakers				
Raw RT	539 (175)	610 (211)	579 (235)	577 (152)
Log RT	6.25 (0.27)	6.37 (0.28)	6.31 (0.28)	6.33 (0.23)
Non-native speakers				
Raw RT	707 (243)	716 (312)	653 (212)	729 (259)
Log RT	6.52 (0.29)	6.51 (0.33)	6.44 (0.28)	6.54 (0.31)

from 17 non-native speakers and 19 native speakers. The non-word data and filler items were not included in the analysis. Incorrect responses, which constituted 2% of the data for both native and non-native speakers, were removed. Extreme values (RTs longer than 3000ms) were also removed; for both native speakers and non-native speakers this represented less than 1% of the data.

The non-native speaker results were then adjusted to take into account any unknown vocabulary items, which removed 17% of the non-native speaker data. The distribution of unknown words was comparable for each of the conditions (Chinese idioms = 22 unknown words, Chinese controls = 21; English idioms = 36, English controls = 31).¹

There were no significant differences in terms of errors for either native or non-native speakers (native speakers ANOVA by condition, $p = .74$; non-native speakers, $p = .98$). Only correct RTs were submitted to further analysis. Unsurprisingly, native speakers had shorter RTs overall than non-native speakers (NS mean = 576 ms, S.D. = 197; NNS mean = 701 ms, S.D. = 259), and independent samples t -tests showed that the difference was significant: $t_1(34) = -3.17$, $p < .01$; $t_2(45.6) = -8.23$, $p < .001$. Patterns of performance for each group were analysed separately with linear mixed effects models using R (R Development Core Team, 2009) and the lme4 (Bates & Maechler, 2009) and languageR packages (Baayen, 2009). Within the models MCMC sampling was used to calculate p -values of all factors.² RTs were log-transformed to reduce skewing as far as possible and Log RT was taken

as the dependent variable. Distribution of RTs for both participant groups is shown in Table 3.

Native speakers

A linear mixed effects model was fitted with the original language of each phrase (English vs. Chinese) and phrase type (idiom vs. control) as fixed effects. List, target length and log-transformed target frequency were also included as fixed effects, as were the experimental factors of trial order and RT to the preceding item. Subject and item were treated as crossed random effects. Non-significant effects were removed from the model (list, $p = .52$; target length, $p = .35$; log-transformed target frequency, $p = .21$).

There was a significant effect of language ($\beta = 0.0630$; $t(738) = 2.53$; $p < .01$) and phrase type ($\beta = 0.1166$; $t(738) = 4.69$; $p < .001$). The interaction between the two was also significant ($\beta = -0.1008$; $t(738) = -2.88$ $p < .01$). These effects were confirmed by fitting separate mixed effects models for the English and Chinese stimuli. For English idioms vs. controls, phrase type was significant ($\beta = 0.1159$; $t(367) = 4.30$; $p < .0001$), while for Chinese idioms vs. controls it was not ($\beta = 0.0170$; $t(369) = 0.75$; $p = .44$).

Non-native speakers

The vocabulary adjusted values were used for analysis of the non-native speaker data.³ A linear mixed effects model was fitted to assess the effects of original language and phrase type. Fixed and random effects were the same as for the non-native speakers. Non-significant effects were removed (list, $p = .69$).

¹ The larger number of errors for English idioms probably reflects the number of words that are commonly used in English in an idiomatic sense only; non-natives are therefore unlikely to encounter such items in literal contexts or in isolation. For example, no non-native-speaker correctly identified the definition of *buck* as used in the English idiom *pass the buck*.

² There is an ongoing debate about how to calculate degrees of freedom, and therefore significance values, in mixed effects modelling. In this paper the degrees of freedom are calculated as ((number of observations) – (number of fixed effects)). The p -values of any effects are estimated through the MCMC sampling process.

³ The non-vocabulary adjusted values were also analysed and a comparable pattern of results was found: Chinese idioms (mean = 668 ms) were responded to significantly faster than Chinese controls (mean = 761ms). This was confirmed using a mixed effects model, where the interaction between language and phrase type was significant ($\beta = 0.0996$; $t(659) = 2.08$; $p < .05$); the effect of phrase type was significant for Chinese ($\beta = 0.0630$; $t(326) = 1.90$; $p < .05$) but not English ($\beta = -0.0520$; $t(329) = -1.77$; $p = .10$) stimuli.

Neither language ($\beta = -0.0440$; $t(550) = -1.25$; $p = .16$) nor phrase type ($\beta = -0.0271$; $t(550) = -0.76$; $p = .40$) were significant on their own but their interaction did approach significance ($\beta = 0.0796$; $t(550) = 1.62$; $p = .07$). To explore this further, separate linear mixed effects models were fitted for Chinese phrases and English phrases. Phrase type (idiom vs. control) was significant for Chinese phrases ($\beta = 0.0664$; $t(285) = 1.93$; $p < .05$) but not English phrases ($\beta = -0.0339$; $t(261) = -1.04$; $p = .31$).

Familiarity, compositionality and proficiency

Because it has been suggested that familiarity (Tabossi et al., 2009; Van Lancker Sidtis, 2012) and compositionality (Gibbs et al., 1989; Gibbs, 1991; Caillies & Butcher, 2007) influence idiom processing, these factors were explored further using linear mixed effects models.

All idioms were very familiar to their native speaker groups and relatively unfamiliar to the opposite groups (on a seven-point Likert scale, on which 1 is completely unfamiliar and 7 is highly familiar, English idioms for native speakers = 6.4 and for Chinese native speakers = 2.8; Chinese idioms for Chinese native speakers = 6.5 and for English native speakers = 2.8). For native speakers relative familiarity was not a significant variable for English idioms ($\beta = -0.0146$; $t(182) = -1.22$; $p = .24$), but it was marginally significant for Chinese idioms ($\beta = -0.0170$; $t(185) = -1.83$; $p = .09$). This suggests that the English items, being at or near a ceiling of familiarity, showed very little variation in RTs according to fine-grained differences. The Chinese items that are more predictable seem to have been judged as more familiar, for example, *doesn't know good from bad* was judged as familiar by English native speakers (mean = 6.4/7), even though it is not a common English phrase (0 occurrences in the BNC). Inclusion of association norms taken from the Edinburgh Associative Thesaurus (Kiss, Armstrong, Milroy & Piper, 1973) confirms this (i.e. using the score for the association between *good* and *bad* as an index of predictability): including this variable as a fixed effect was significant ($\beta = -0.1580$; $t(184) = -2.11$; $p < .05$), and this removed the effect of familiarity for Chinese idioms ($\beta = -0.0098$; $t(184) = -1.00$; $p = .34$). Non-native speakers showed no variation according to how relatively familiar the idioms were: familiarity was not significant for English idioms ($\beta = -0.0002$; $t(130) = -0.02$; $p = .98$) or Chinese idioms ($\beta = -0.0171$; $t(141) = -0.96$; $p = .42$). Taken together these results indicate that relative familiarity did not modulate RTs for idiom completions, but it should be remembered that all items were deliberately chosen to be highly familiar, so this lack of variation is perhaps unsurprising.

Compositionality was also included in the analysis to assess its contribution to RTs. Two measures were used: compositionality ratings from English native speakers (judgement of English forms of both English and Chinese idioms) and an additional rating of the Chinese idioms in the original Chinese characters by a set of Chinese native speakers. We assumed that all control items are potentially just as compositional as their corresponding idioms, i.e. for native English speakers, the Chinese idiom *draw a snake and add feet* and the control *draw a snake and add hair* could both just as easily mean “ruin with unnecessary detail”, hence they are equally compositional. In addition, because the prime phrases are the same (e.g. *draw a snake and add . . .*), the contribution of the compositionality of the prime phrase must be comparable across the idiomatic and control conditions. Table 4 summarises the results of analysis according to compositionality, showing analysis of all stimuli (idioms and controls) and of the idiom conditions separately.

For native English speakers none of the measures of compositionality demonstrated an influence on RTs for either set of stimuli. For non-native speakers the English idioms were not affected by compositionality, and for the Chinese items only the English native speaker judgements of the translated versions were significant. When idioms and controls were considered together there was a marginally significant effect of compositionality ($\beta = 0.0274$; $t(284) = 1.86$; $p = .07$) and a significant interaction with phrase type ($\beta = -0.0520$; $t(284) = -2.40$; $p < .05$). Analysis of the conditions separately showed a significant effect of compositionality for idioms ($\beta = 0.0303$; $t(141) = 2.19$; $p < .05$) but not controls ($\beta = -0.0234$; $t(139) = -1.14$; $p = .23$), so it is clear that compositionality did not affect the advantage for the idioms over controls for non-native speakers (when compositionality was included the difference between idioms and controls was still significant: $\beta = 0.2608$; $t(284) = 3.02$; $p < .01$), but the idioms themselves were affected by the degree of compositionality. This supports Caillies and Butcher (2007), who found an advantage for decomposable over non-decomposable idioms, but it should be noted that their study looked at meaning activation (lexical decision task on targets related to the figurative meaning), which was not required of the participants in our study. For future studies it will be important to define and control the dimension of compositionality very carefully to establish its exact role in cross-language idiom processing, especially if studies are concerned with recognition/activation of form as opposed to meaning.

Proficiency level can also play a role in non-native idiom processing (Conklin & Schmitt, 2008; Ueno, 2009), so this was explored as a factor for the non-native speakers. Non-native speaker proficiency based on vocabulary score was non-significant for overall performance ($\beta =$

Table 4. *Contribution of compositionality to response times to English and Chinese stimuli. Values are MCMC-estimated p-values based on t-scores in linear mixed effects models with compositionality rating included as a fixed effect.*

	English stimuli English judgements	English judgements	Chinese stimuli Chinese judgements
Idioms + controls			
Native speakers	.95	.52	.88
Non-native speakers	.61	.07	.82
Idioms only			
Native speakers	.92	.44	.91
Non-native speakers	.24	.04*	.46

* $p < .05$

−0.0107; $t(549) = 0.42$; $p = .66$) or as part of a three-way interaction with language and phrase type ($\beta = 0.0076$; $t(549) = 0.29$; $p = .80$). All other direct measurements of proficiency (vocabulary score, self-ratings of speaking, reading, writing and listening skills and usage score) were shown to be non-significant (all p -values $> .05$). The only significant indicator was the length of time studying English ($\beta = 0.0313$; $t(549) = -1.67$; $p < .05$), which may simply show that longer exposure leads to a better ability to recognise and judge English words (greater lexical knowledge, awareness of English forms, etc.). Importantly, analysis of the English and Chinese materials separately showed no interaction with phrase type for English items ($\beta = -0.0086$; $t(261) = -0.85$; $p = 0.39$) or Chinese items ($\beta = 0.0044$; $t(284) = 0.45$; $p = .68$). Longer exposure to English therefore improved RTs across the board, but did not affect the pattern of performance for any participant.

The lack of any direct effect of proficiency may be unsurprising given the homogenous nature of the non-native participant group. All were from the same study-abroad cohort and had broadly comparable proficiency and experience in English. In contrast, Ueno (2009) manipulated proficiency and found a significant difference between high and low proficiency groups. It is likely that in order to see an influence of proficiency, we would need to look at participants with a wider range of proficiencies.

Discussion

A clear pattern of results for native and non-native speakers was observed. Both native and non-native speakers responded most quickly to targets that formed idioms in their respective L1s, and the difference relative to matched control items was significant. The native speaker results are important as they support multiple previous studies showing an advantage for idioms over matched novel language. They also show a clear pattern of performance according to overall familiarity: the English

idioms showed an advantage over control items because they were known, whereas the Chinese idioms were not, so RTs in the idiom and control conditions did not differ. Importantly, the English idioms showed no grading according to familiarity, so more familiar idioms were not significantly faster than less familiar ones. This may be simply be a reflection of the fact that stimuli were deliberately chosen to be common and familiar, so any variation was likely to be extremely fine-grained (probably too fine-grained to significantly affect the RTs). Compositionality was not a significant factor either for fundamentally familiar (English) or unfamiliar (Chinese) idioms. The native speaker results support the assertions of Tabossi et al. (2009) and Van Lancker Sidtis (2012) that overall familiarity (whether an item was known or unknown) is the main driver of idiom recognition and therefore formulaicity.

A complementary pattern of results was observed for non-native speakers. Targets that formed English idioms were not reliably faster than controls, suggesting that these had not been encountered with enough regularity to form phrasal representations in English, which is contrary to evidence that advanced non-natives show a formulaic advantage (Isobe, 2011; Jiang & Nekrasova, 2007; Underwood, Schmitt & Galpin, 2004). This is, however, in line with the general inconsistency of results, in which non-native speakers sometimes show a processing advantage and other times do not. Chinese idioms, despite being presented in an entirely unfamiliar form (English), did show an advantage over the control phrases. Relative familiarity within the idiom condition was not significant, suggesting that it was simply the status as known (idioms) or unknown (controls) that drove the advantage.

The finding that RTs to translations of L1 idioms by Chinese speakers are shorter poses an interesting problem for the dual route model. Van Lancker Sidtis (2012) suggested that formulaic expressions differ from other utterances because they are not newly

created. Importantly, in a purely formal/lexical sense, the translated Chinese idioms were novel, and the non-native participants are highly unlikely to have encountered the sequences in English (as evidenced by their 0 frequency in the BNC and the lack of familiarity for native speaker participants). Thus a canonical, learned configuration, stored as a result of many previous encounters and activated via associative lexical links, cannot explain the advantage observed for the translated idioms. What therefore accounts for the advantage for the translated Chinese idioms, and can this advantage be explained by the dual route model?

One possible explanation is that idioms can be activated at a conceptual level. Unlike some other forms of formulaic language, idioms have their own separate conceptual entry (i.e. *spill the beans* means *REVEAL A SECRET*); Wray (2012) suggested that it may be this property that offers them an advantage over non-idioms. One view of the bilingual lexicon is that there is an underlying shared conceptual system, so learning L2 items involves the mapping of new forms onto existing concepts. Over time and as proficiency increases, direct links from L2 forms to concepts can be created, allowing bilinguals to bypass the L1 forms (Kroll & Stewart, 1994; Wang, 2007). If this is correct, idioms may exist as unitary concepts that are accessible via lexical forms in either language. Encountering the English prime (e.g. *draw a snake and add...*) therefore activates the underlying concepts of the component words (*DRAW*, *SNAKE*, *ADD*) in the shared bilingual conceptual store, and the associations of these at a conceptual level trigger the idiom entry (*RUIN WITH UNNECESSARY DETAIL*). This unitary concept activates not only the figurative meaning but also the whole phrase and therefore the expected completion (*FEET*), making the lexical form of the target available either directly in the L2 if a strong enough link has been created (e.g. *feet*), or in the L1 (足). Because this L1 form is a translation equivalent of the presented target, facilitation for the English form *feet* is still observed in either case.

Such a view is broadly in accord with the conclusions reached by Ueno (2009) and Wolter and Gyllstad (2011). In their studies of collocations they proposed that lexical forms in the L2 (English) activated associative links in a language non-selective way, i.e. at a conceptual level, so words that would form collocations in the L1 will be primed even when they are encountered in the L2. In particular, Ueno (2009) found that the effect increased with L2 proficiency: her participants' responses to both translated L1 collocations and L2 collocations became shorter as proficiency increased, which she suggested was evidence of a strengthening of the separate links between the L1 and L2 lexical systems and the shared conceptual system. Our results show no variation according to proficiency but do show shorter RTs as a result of increased number of years studying English. This may

suggest that increased exposure can lead to more efficient access to L2 forms (or possibly just better ability to judge English words/non-words), but without a more rigidly defined set of high and low proficiency participants it is difficult to say any more about the development of direct conceptual access. If Ueno's hypothesis is correct, we would expect a higher proficiency group to show a more pronounced idiom superiority effect for the translated idioms, and probably also an effect for English idioms, as increased exposure would be likely to generate idiom entries, at least for the most frequent English items.

A conceptual basis for cross-language priming beyond the single word level is therefore plausible, but our results do not provide unequivocal support for this. The task was designed to investigate whether the form of an idiom was the principle driver of recognition; participants therefore did not need to access any conceptual information in order to complete the task, because a lexical decision could be based solely on the form of the target word rather than on any associated semantic meanings (literal or figurative). A lexical translation-based process may therefore provide an alternative way to account for the results.

Zhang, van Heuven and Conklin (2011) demonstrated the process of fast automatic translation for Chinese–English bilinguals. They used English word pairs in a masked priming task with very short presentations (59 ms) and found that the Chinese translation of the prime word was influential (i.e. when the prime-target showed a repeated morpheme in the Chinese translation there was facilitation which was not present when the prime-target produced translations with unrelated morphemes). They concluded that the participants must be translating and decomposing the English primes quickly and automatically for the Chinese morphology to show an effect in a completely English task. The same process may be at work in the current study. Presentation of the prime phrases could be quickly and automatically translated and decomposed, with the result that the L1 characters are activated and their associations as part of an idiom are recognised at an L1 lexical level. This activates the overall Chinese idiom, which primes the final character; because this is a translation equivalent of the target in English, facilitation for the L2 form is observed.

In this explanation the configuration priming the idiom is language-specific in that it is driven by associations at a lexical level in Chinese. Wang (2007) showed inter-language priming only for direct translation equivalents; in the current study, whilst the individual words are translation equivalents, the phrases are not (they do not exist in both languages), so any associations at a lexical level must be driven by the L1 (Chinese). Wang (2007) highlights another important factor: the influence of strategic processes. In the current study primes were not masked and were presented in a self-timed protocol, potentially giving participants ample time to read and

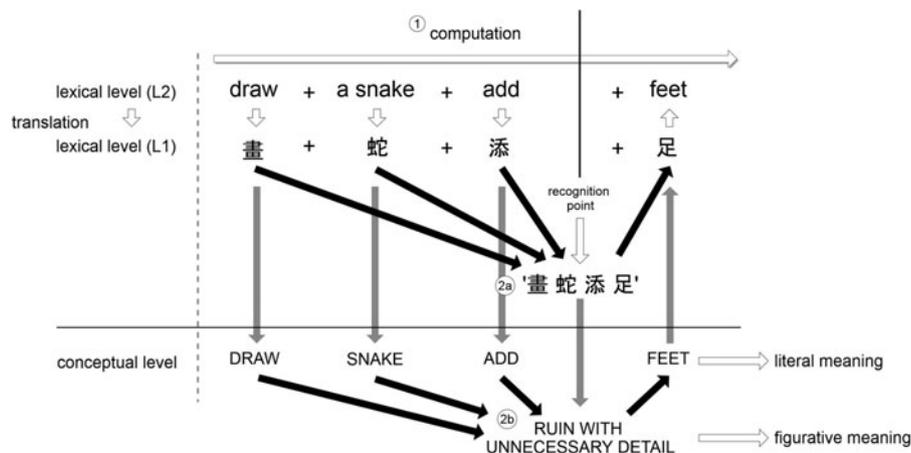


Figure 2. Modified dual route model for the translated idiom *draw a snake and add feet*. In this model two routes are available: analysis and computation of the phrase (1), and direct access either via a translation-based route at the lexical level (2a) or via a conceptual route (2b). In both of the direct routes a unitary entry is accessible, either as a lexical configuration (2a) or a distinct underlying concept (2b). Black arrows represent associative links between components, white arrows represent processes and grey arrows represent links between lexical items and their underlying concepts.

translate them, make associations in the L1 and predict the final character, which would yield shorter RTs when the English target was a translation equivalent of the expected completion. Idioms present a particular challenge in this regard because their length means they are generally unsuitable for masked priming, so alternative methods may be required in future to disambiguate automatic and strategic translation processes for such stimuli. However, whether translation was fast and automatic or strategic, an influence of the known L1 configurations was still observed in the L2. As with the conceptual explanation, increased proficiency might affect the process: as the lexical links between L1 and L2 are reinforced, activation of L1 forms via the L2 would become faster, so the idiom advantage might also become stronger if the effect is driven by lexical/translation processes.

Proposing that faster processing for L1 configurations in an L2 could have a conceptual or lexical basis broadly reflects the distinction made by Bley-Vroman (2002), who identified both a lexical frequency-based and a meaning-based motivation for formulaic language processing. Both explanations for our results can be incorporated into a dual route model, as shown in Figure 2.

The modified dual route model allows bilinguals to access L1 idioms even when they are presented in the L2. An important consideration is how non-natives have been shown to process formulaic language in the L2. With idioms in particular, Cieřlicka (2006) suggests that there is a fundamental difference in approach for native and non-native speakers: broadly speaking, native speakers tend to use a retrieval route wherever possible whereas non-natives are more likely to approach all material compositionally. In our results no difference between English idioms and controls was observed for non-native

speakers, which suggests that the default approach for both sets of participants was to process the idioms compositionally. The difference between the groups is that for native speakers an additional configuration was recognised and unlocked, whereas for non-native speakers no such direct route was available. This indicates not necessarily a difference in approach, but rather a difference in available resources: non-native speakers are less likely to have formed associative links that can unlock the lexical configuration of an idiom and its underlying concept. Matlock and Heredia (2002) suggested that this leads to a situation in which non-native speakers only recognise phrases as idioms once they have analysed them and found them to be incongruent.

For Chinese speakers encountering English idioms, even if they are recognised as non-compositional configurations and are potentially easy to “spot” as idioms, no underlying lexical or conceptual configuration may be available. The Chinese idioms presented in English did show an effect of compositionality for the Chinese speakers if we take the compositionality ratings from English native speakers, but this did not negate the advantage they have over control phrases (if we assume that the control phrases are as compositional as their corresponding idioms). Similarly, taking the potentially more meaningful Chinese ratings of compositionality, all effects of this variable are non-significant for the Chinese idioms. This is consistent with the findings of Tabossi et al. (2009), who showed an overall advantage for familiar phrases but no variation for compositional items (clichés) compared with non-compositional items (idioms). Results from other studies in this respect have been mixed (for example, Gibbs et al., 1989; Gibbs, 1991; Caillies & Butcher, 2007), but a reasonable conclusion

seems to be that the compositionality is strongly linked to meaningfulness and familiarity: the Chinese native speakers' ratings suggest that because the phrases (and their underlying stories) were known, the process of mapping idiomatic meaning onto the lexical items was facilitated. Including the Chinese speakers' ratings in the analysis, rather than the potentially "purer" but less meaningful English ratings, demonstrated that the overall contribution of compositionality was not significant for the present task, i.e. form-based recognition.

For native speakers no effect of compositionality was observed for either set of idioms. For the English idioms, this is in line with some previous research (e.g. Tabossi et al., 2009). For translated Chinese idioms, compositionality did not affect native speaker processing; this is unsurprising because no lexical or conceptual configurations would be available to aid English speakers' recognition for any of the Chinese idioms. Again, these results support Tabossi et al. (2009) rather than, for example, Caillies and Butcher (2007), in implicating overall idiom familiarity (known or unknown) as the key driver of the idiom superiority effect. Thus, English idioms, which are familiar and well known, show a processing benefit, while Chinese idioms, which are unfamiliar and unknown, are processed at the same speed as control items, but the degree of compositionality does not significantly affect either set of items. One important caveat to these results is that all idioms – both English and Chinese – were deliberately chosen to be highly familiar and the degree of compositionality was not controlled in advance, i.e. no deliberate contrast of "high" vs. "low" compositionality was adopted as in other studies of idiom processing. Both of these complex and multi-faceted variables should be carefully considered in any future studies looking at cross-language idiom processing, to allow researchers to fully investigate their effects.

In conclusion, non-native speakers were shown to respond more quickly to idioms translated from their L1 than to control phrases in a lexical decision task. This result mirrors native speaker performance for English idioms, suggesting that a dual route model can explain bilingual performance as well as monolingual access to formulaic language. Overall familiarity with the L1 form – recognising a "known" phrase – was the main driver of the processing advantage for both native and non-native speakers. The "retrieval" branch of the dual route model for bilinguals may represent a process at the lexical level, in which English items were translated into their Chinese equivalents. This activates a known L1 lexical configuration, facilitating subsequent processing for translation equivalents in the L2. Alternatively, the same associations may exist at a language non-specific conceptual level, suggesting that it is the separate conceptual entry for idioms that drives their processing advantage. While the current results do not allow us to

distinguish between these two explanations, some level of L1–L2 interaction is clearly indicated. This adds further support to the argument that idioms are not represented as single, unanalysable units in the lexicon, but instead represent a distributed meaning that is accessed via the component words.

References

- Baayen, R. H. (2009). languageR: Data sets and functions with "Analysing linguistic data: A practical introduction to statistics". R Package version 0.955.
- Bates, D. M., & Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32.
- Bley-Vroman, R. (2002). Frequency in production, comprehension, and acquisition. *Studies in Second Language Acquisition*, 24, 209–213.
- Bobrow, S., & Bell, S. (1973). On catching on to idiomatic expressions. *Memory and Cognition*, 1, 343–346.
- Cacciari, C., & Glucksberg, S. (1991). Understanding idiomatic expressions: The contribution of word meanings. In G. Simpson (ed.), *Understanding word and sentence*, pp. 217–240. Amsterdam: Elsevier.
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27, 668–683.
- Caillies, S., & Butcher, K. (2007). Processing of idiomatic expressions: Evidence for a new hybrid view. *Metaphor and Symbol*, 22, 79–108.
- Chen, H., & Ng, M. (1989). Semantic facilitation and translation priming effects in Chinese–English bilinguals. *Memory and Cognition*, 17, 454–462.
- Chung, K., Code, C., & Ball, M. (2004). Lexical and non-lexical speech automatisms in aphasic Cantonese speakers. *Journal of Multilingual Communication Disorders*, 2, 32–42.
- Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, 22, 115–144.
- Code, C. (1994). Speech automatization production in aphasia. *Journal of Neurolinguistics*, 8, 135–148.
- Code, C. (2005). First in, last out? The evolution of aphasic lexical speech automatisms to agrammatism and the evolution of human communication. *Interaction Studies*, 6, 311–334.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29, 72–89.
- Cutting, J., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition*, 25, 57–71.
- De Groot, A., & Nas, G. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30, 90–123.
- Durrant, P. (2008). High frequency collocations and second language learning. Ph.D. dissertation, University of Nottingham.

- Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly*, 42, 375–396.
- Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20, 29–62.
- Foster, P. (2001). Rules and routines: A consideration of the role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan & M. Swain (eds.), *Researching pedagogic tasks: Second language learning, teaching and testing*, pp. 75–95. Harlow, UK: Longman.
- Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory and Cognition*, 8, 149–156.
- Gibbs, R. W. (1991). Semantic analyzability in children's understanding of idioms. *Journal of Speech and Hearing Research*, 34, 613–620.
- Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, 28, 576–593.
- Google Translate. <http://translate.google.co.uk/>.
- Grant, L., & Bauer, L. (2004). Criteria for re-defining idioms: Are we barking up the wrong tree? *Applied Linguistics*, 25, 38–61.
- Isobe, Y. (2011). Representation and processing of formulaic sequences in L2 mental lexicon: How do Japanese EFL learners process multi-word expressions? *JACET Kansai Journal*, 13, 38–49.
- Jiang, N., & Nekrasova, T. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91, 433–445.
- Kiss, G. R., Armstrong, C., Milroy, R., & Piper, J. (1973). An associative thesaurus of English and its computer analysis. In A. Aitken, R. Bailey & N. Hamilton-Smith (eds.), *The computer and literary studies*. Edinburgh: Edinburgh University Press.
- Kroll, J., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, 33, 149–174.
- Kuiper, K., van Egmond, M., Kempen, G., & Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon*, 2, 313–357.
- Lin, M., & Leonard, S. (2012). *Dictionary of 1000 Chinese idioms* (revised edition). New York: Hippocrene Books.
- Liu, Y., Li, P., Shu, H., Zhang, Q., & Chen, L. (2010). Structure and meaning in Chinese: An ERP study of idioms. *Journal of Neurolinguistics*, 23, 615–630.
- Matlock, T., & Heredia, R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. Heredia & J. Altarriba (eds.), *Bilingual sentence processing*, pp. 251–274. Amsterdam: Elsevier.
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31, 9–13.
- Omazic, M. (2008). Processing of idioms and idiom modifications. In S. Granger & F. Meunier (eds.), *Phraseology: An interdisciplinary perspective*, pp. 67–79. Amsterdam: John Benjamins.
- On-line Chinese Tools Chinese/English Dictionary. Available at <http://www.mandarintools.com/worddict.html>.
- 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. 191–225. London: Longman.
- R Development Core Team. (2009). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.
- Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *Quarterly Journal of Experimental Psychology*, 55A, 1339–1362.
- Schmitt, N. (2005). Formulaic language: Fixed and varied. *Estudios de Lingüística Inglesa Aplicada*, 6, 13–39.
- Simon, H., Zhang, W., Zang, W., & Peng, R. (1989). STM capacity for Chinese words and idioms with visual and auditory presentations. In H. Simon (ed.), *Models of thought* (vol. 2), pp. 68–75. New Haven, CT: Yale University Press.
- Siyanova, A. (2010). On-line processing of multi-word sequences in a first and second language: Evidence from eye-tracking and ERP. Ph.D. dissertation, University of Nottingham.
- Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54, 161–184.
- Swinney, D., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behaviour*, 18, 523–534.
- Tabossi, P., Fanari, R., & Wolf, K. (2008). Processing idiomatic expressions: Effects of semantic compositionality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 313–327.
- Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory and Cognition*, 37, 529–540.
- Tremblay, A., & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioural and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (ed.), *Perspectives on formulaic language: Acquisition and communication*, pp. 151–173. London: Continuum.
- Ueno, T. (2009). An investigation of the relationship between the development of bilingual semantic organisation and interactive connectivity across languages. Ph.D. dissertation, Trinity College, Dublin.
- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (ed.), *Formulaic sequences*, pp. 153–172. Amsterdam: John Benjamins.
- Van Lancker Sidtis, D. (2012). Two-track mind: Formulaic and novel language support a dual-process model. In M. Faust (ed.), *The handbook of the neuropsychology of language*, pp. 342–367. Chichester, UK: Wiley-Blackwell.
- Van Lancker Sidtis, D., & Postman, W. A. (2006). Formulaic expressions in spontaneous speech of left- and right-hemisphere-damaged subjects. *Aphasiology*, 20, 411–426.
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22, 1682–1700.

- Wang, X. (2007). Does semantic activation spread across languages? An experimental study with Chinese–English bilinguals. *Coyote Papers*, 15, 90–99.
- Warren, H. (1994). *Oxford learner's dictionary of English idioms*. Oxford: Oxford University Press.
- Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics*, 32, 430–449.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.
- Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford: Oxford University Press.
- Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, 32, 231–254.
- Wray, A., & Perkins, M. (2000). The functions of formulaic language: An integrated model. *Language and Communication*, 20, 1–28.
- Zhang, H., Yang, Y., Gu, J., & Ji, F. (2013). ERP correlates of compositionality in Chinese idiom comprehension. *Journal of Neurolinguistics*, 26, 89–112.
- Zhang, T., van Heuven, W. J., & Conklin, K. (2011). Fast automatic translation and morphological decomposition in Chinese–English bilinguals. *Psychological Science*, 22, 1237–1242.
- Zhou, S., Zhou, W., & Chen, X. (2004). Spatiotemporal analysis of ERP during Chinese idiom comprehension. *Brain Topography*, 17, 27–37.