CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

ORCA – Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:https://orca.cardiff.ac.uk/id/eprint/84266/

This is the author's version of a work that was submitted to / accepted for publication.

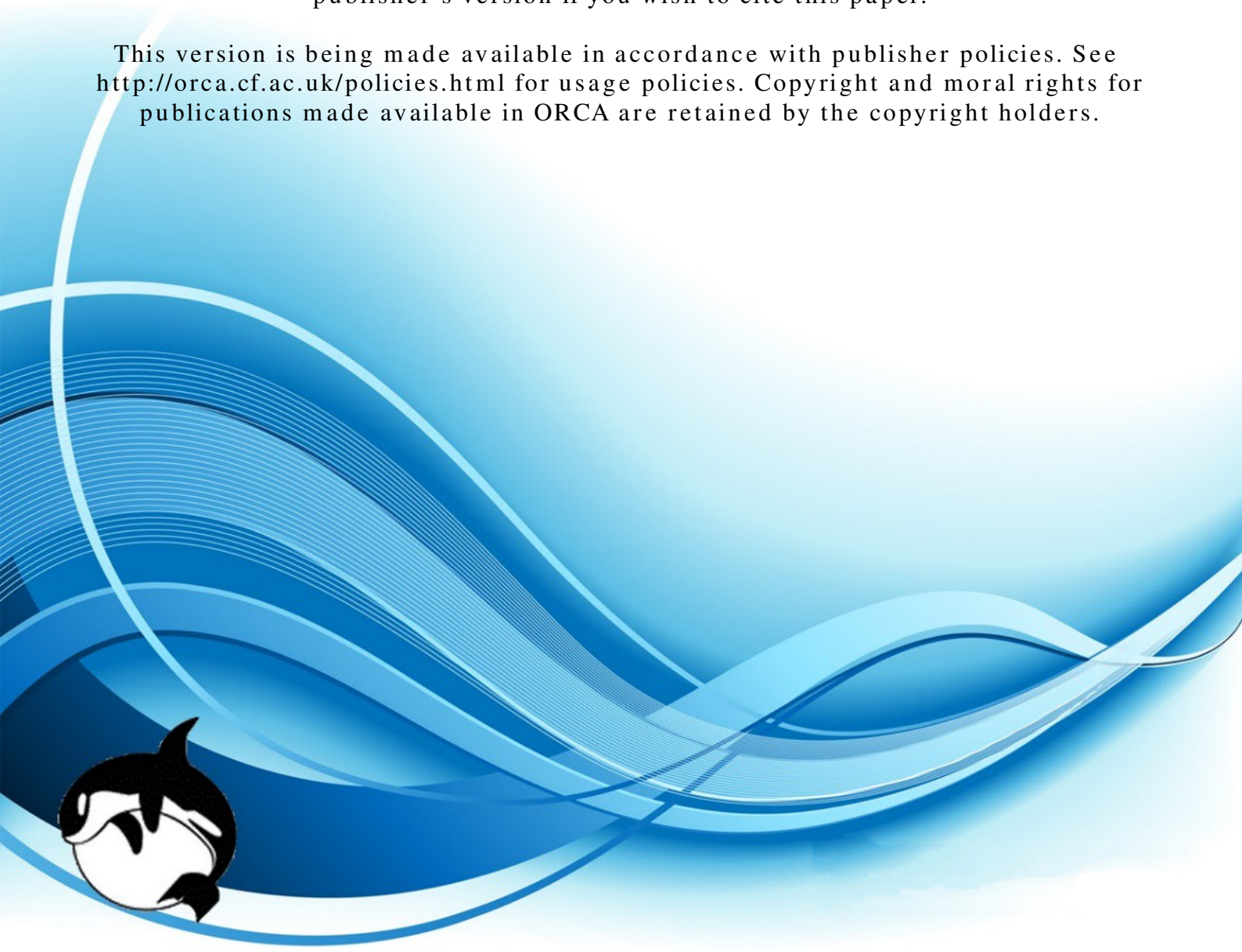
Citation for final published version:

Wray, Alison , Bell, Huw and Jones, Katy 2016. How native and non-native speakers of English interpret unfamiliar formulaic sequences. European Journal of English Studies 20 (1) , pp. 47-63. 10.1080/13825577.2015.1136163

Publishers page: http://dx.doi.org/10.1080/13825577.2015.1136163

Please note:
Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See http://orca.cf.ac.uk/policies.html for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



How native and non-native speakers interpret unfamiliar formulaic sequences

Alison Wray (corresponding author)
Centre for Language and Communication Research
School of English, Communication and Philosophy
Cardiff University
John Percival Building
Colum Drive
Cardiff, CF10 3EU
wraya@cardiff.ac.uk

Alison Wray is a Research Professor in Language and Communication at Cardiff University. She holds a BA and D.Phil from the University of York, UK. Her main contribution to research has been in the domain of formulaic language, including two monographs (2002 and 2008), the first of which won the 2003 book prize of the British Association for Applied Linguistics.

Huw Bell
Language, Linguistics and TESOL
Manchester Metropolitan University
227 Mabel Tylecote Building
Cavendish Street
Manchester, M15 6BG
h.bell@mmu.ac.uk

Huw Bell is a Senior Lecturer and researcher in applied linguistics at Manchester Metropolitan University. His PhD is on vocabulary acquisition, from Swansea University, and he has an RSA Diploma in TEFL. He is currently working on a project examining the teaching of formal grammar in primary schools. Other research interests include the L1 and L2 lexicon.

Katy Jones
Centre for Language and Communication Research
School of English, Communication and Philosophy
Cardiff University
John Percival Building
Colum Drive
Cardiff, CF10 3EU
jonesKS1@cardiff.ac.uk

Katy Jones is a researcher and part time lecturer in Language and Communication at Cardiff University. She holds an MA from the University of Sussex. She taught EFL in the UK, Spain and Japan and was a tutor of English for Academic Purposes at Cardiff University before beginning her PhD there. Her thesis (awarded 2014) was on referring expressions. She is currently involved in several research projects, including investigating the role of the noun phrase in (digital) language development, examining the exploitation of (in)definiteness in persuasive discourse, and comparing the advice about communication that is given to carers of people with dementia.

Word count (excluding title page, tables, appendices and figures, but including abstract and references): **7115**

How native and non-native speakers of English interpret unfamiliar formulaic sequences

Alison Wray (Cardiff University)
Huw Bell (Manchester Metropolitan University)
Katy Jones (Cardiff University)

Abstract
This study examines whether native and non-native speakers faced with unfamiliar formulaic expressions use the same tactics for working out what they mean. The test items needed to be semantically opaque, used in an authentic context, and unknown to all participants. Ten obsolete expressions were selected from the historical novels of Georgette Heyer. First language English speakers and UK-resident classroom-taught learners of English as a foreign language were individually presented with the expressions in their original context, and asked to work out what they meant. Analysis of their comments revealed that the native speakers deployed significantly more context and analogy. The non-native speakers were much more likely than native speakers to refer to individual unknown words. Whilst it seems that first language users take a more holistic approach to linguistic input than classroom-taught second language learners do, the findings may suggest that learners adopt increasingly 'native-like' strategies as proficiency increases.

Key words: formulaic language; idioms; language learning; comprehension; analogy; context

1. Introduction

MAGGIE	The wife of a board member died and Will asked me to send flowers. I wrote on the card, 'I'm so sorry about your loss. LOL.'
JIM	LOL?
MAGGIE	I thought it meant 'lots of love'.

(*The Newsroom*, Series 1, Episode 6, HBO, 2012)

When we first encounter a new expression, how do we work out what it means? If it is not directly analysable, as with *LOL*, we will either have to ask someone, or make a guess. A guess will be based on two possible types of information – internal cues (what could the Ls and the O reasonably stand for in this context?) and external cues (what do others who use this expression appear to mean?). Sometimes, neither type of cue is sufficient, and, just as with single words, individuals can come to different conclusions that may remain undetected until some incident exposes a misunderstanding.[1]

The same can happen when the components of the expression are more transparent. A person unfamiliar with *kick over the traces*, and seeing it in the sentence in Example 1 would have to use context to work out the likely meaning.

Ex.1
At the age of sixty, Walter *kicked over the traces* and ran away to Brazil[2]

[1] LOL is defined by the OED and American Heritage Dictionary as meaning 'laugh(ing) out loud' but other interpretations, including 'lots of love' are also attested, most famously in the usage of UK Prime Minister David Cameron, http://www.businessinsider.com/david-cameron-lol-lots-of-love-rebekah-brooks-2012-5

[2] Example taken from http://idioms.thefreedictionary.com/kick+over+the+traces

Walter did something (ran away) in the latter part of his life, and the deed was tantamount to kicking over the traces. So, what might it mean? The main lexical indicators are *kick* and *traces*. Taking *kick over* to have an abstract meaning founded on an image of physical action, several options are possible: kicking a ball over a fence or over to a person, kicking one's foot over a high object, or covering something up by kicking dust, leaves or other objects over it. To understand which sort of *kick over* it is, we need to understand what *traces* means. There are two options. The more common meaning is *marks, signs* or *vestiges* as in *there were no traces of a struggle.* But there is another meaning, which the Oxford English Dictionary (OED) describes as 'obsolete': *The pair of … leather straps by which the collar of a draught-animal is connected with the splinter-bar … .* This is the meaning that applies in the idiom: *to get a leg over the traces so as to kick more freely and vigorously;* fig. *to throw off the usual restraints* (OED).

Clearly, a problem will arise in understanding the metaphorical meaning of *kick over the traces* if the second meaning of *traces* is not known or is not chosen. Nor will it necessarily be obvious what has happened, since the more common meaning of *traces* can also offer a plausible meaning for the expression. Someone who does not want to leave any footprints might well kick dust over them. Perhaps, then, when Walter *kicked over the traces* he covered up his tracks. He made himself disappear when he went to Brazil, so that no one could follow him. Two people could potentially interact successfully for some time with different interpretations of this idiom before anything in the usage or context revealed the non-alignment.

This example shows how when we attempt to infer the meaning of unfamiliar, non-literal expressions in use, we have to manage a range of uncertainties, ambiguities and potentially conflicting information. Mistakenly choosing the more common meaning of *traces* is only possible because it renders a plausible alternative interpretation. Thus, context, analogy, precedent and pragmatics play their role alongside frequency. Suppose the more common meaning of *traces* had not rendered a plausible reading, but the infrequent meaning was also not known. How would one then work out its meaning? That is the question addressed in this article.

2 Why idioms are difficult to learn

Idioms differ from other multiword expressions in two main ways. Firstly, they have a form that is recognised within a speech community as being associated with a particular non-literal meaning (Laval & Bernicot, 2002). A useful rule of thumb for judging whether an item is an idiom or not is to ask whether, knowing the meaning, one could possibly come up with the form if one didn't know it already. By this reckoning, *dress someone down, throw the book at someone* and *look a gift horse in the mouth* are idioms, where *dress someone up, throw a ball at someone* and *look a gift salesman in the eye* are not. The reason is that the latter mean what they appear to mean, whereas the former are a conventionalised way of expressing a meaning different from the surface one.

Secondly, their form is fixed – at least to an extent. In fact, although idioms need some core static component, few of them are completely immutable (Moon, 1998). Function words, adverbs and adjectives are likely to be stable, as is the verb (though not its morphology). Sometimes nouns are fixed. But often the expression is kept flexible for different referents through a noun phrase slot (e.g. *NP_1 give NP_2 a piece of NP_1's mind*).

Marilyn Nippold (2006: 370) locates idioms along a continuum of difficulty for learners, from those that are 'transparent and frequently occur in spoken and written language' to those that are 'opaque and rarely occur in the language'. Transparency here does not mean

that the words have their literal meaning, but rather that it is possible to map the elements in the expression onto components of the non-literal meaning. To illustrate, Nippold contrasts *go by the book*, where the 'book' can be imagined to contain the rules that are complied with, with *go by the board*, where the meaning of 'be discarded' cannot be mapped so easily onto the internal components.

Despite Nippold's implication, transparency and frequency are not strongly associated. They are at least independent if not in an inverse relationship. As with individual lexical items, frequency protects idiosyncratic forms and meanings from being regularised, meaning that frequency and irregularity are typically associated (Rumelhart and McClelland, 1986: 240). Low transparency does not inevitably result in greater difficulty with comprehension. Other than in citation contexts, the core arbiter of comprehensibility is a third variable, context. Indeed, the natural learning of idioms requires inference from their usage. Although, as we saw earlier, it is not an infallible process, context will generally trump transparency. If that were not the case, then when we encountered a word string that was ambiguous (e.g. *skating on thin ice*) we would favour the easier, literal reading, even if it did not make sense. It is only context that tells us when such an expression is intended metaphorically rather than literally. The incongruence that arises in metaphorical usage requires the hearer or reader to apply pragmatics to work out a plausible meaning, according to Paul Grice's (1975) principle of cooperation – that is, the assumption that the speaker or writer intended to be meaningful.

Thus, the survival of idioms in a language is dependent on the capacity of hearers and readers to suspend disbelief in relation to the literal meaning and look for a metaphorical meaning. However, we do not have to rerun that process every time, for idioms seem to be rather easy to remember. The initial impact of the image, and the effort taken to tease out the intended meaning, may make them more salient and thus more memorable. Certainly something is defying the typical patterns of learning, whereby frequency of exposure builds up familiarity (Taylor, 2012). Idioms are too infrequent for that to be the explanation. *Take a rain check* occurs just 19 times in the 450 million word Corpus of Contemporary American English (COCA) and *kick over the traces* occurs 5 times – by way of comparison *hegemonic* occurs 1011 times and *ontological* 837 times.

3 How idioms are learned

3.1 Idiom learning in the first language

Idiom learning in the first language needs to be seen in the context of children's approach to language acquisition more generally. It is possible to explain the progression of children's general language acquisition on the basis of the recycling of the multiword strings to which they are exposed, along with an alertness to the relative fixedness of components within them, so that flexibility can be introduced exactly where it is permitted (Peters, 1983). Elena Lieven and Ewa Dąbrowska (e.g. Lieven et al. 2003; Dąbrowska and Lieven 2005) found that an infant's novel expression almost entirely entailed deploying simple operations (embedding and chaining) to reuse previous input from the mother. Alison Wray (2002) links the holistic approach to first language acquisition to the ubiquity, tenacity and functional importance of formulaic material in a language. As Michael Tomasello (2003: 101-102) puts it,

> It turns out that, upon inspection, a major part of human linguistic competence – much more than previously believed – involves the mastery of all kinds of routine formulas, fixed and semi-fixed expressions, idioms, and frozen

> collocations. Indeed one of the distinguishing characteristics of native speakers of a language is their control of these semi-fixed expressions as fluent units with somewhat unpredictable meanings.

Idioms, then, are part of a much larger set of formulaic units in a language. However, there is a difference in how they can be acquired. Tomasello, Lieven and Dąbrowska found their claims on usage-based models of language acquisition, in which frequency of exposure is key:

> Speakers are exquisitely sensitive to the language that they encounter, noting the uses of words, their collocations and their syntactic environments, the constructions and their special semantics and conditions of use, as well. (Taylor, 2012: 283-284).

But, as noted above, idioms are not frequent enough, or flexible enough, to be learned in this way. Mila Vulchanova et al. (2011) found that by age ten, children were still unfamiliar with 50% of the common idioms on which they were tested. Given the semantic impenetrability of idioms, inferential learning from context will only be possible if there are plenty of examples that between them enable the child to identify patterns in their usage.

In lieu of adequate exposure to idioms, children seem to do two things. One is to set them aside for later. As with unknown words, if they are anticipating decoding them on the basis of frequency, then they can only note that they saw them, and must then wait for more information before working out what they mean. The other way children handle idioms is by bringing to bear on them their current knowledge of the language and the world – a world in which figurative language does not yet feature all that strongly. Virginie Laval & Josie Bernicot (2002) found that six-year old children struggled to interpret idiomatic expressions non-literally. However, nine-year olds performed better, by virtue of relying heavily on the context. Meanwhile, adults were able to interpret expressions non-literally even in the absence of context. Vulchanova et al (2011) found that ten-year-olds offered literal readings of unfamiliar idioms – but they were not presented in a context of use. This meant that the children did not have any opportunity to recognise incongruence between a literal meaning of the idiom and how it was used.

If someone unfamiliar with the expression *skate on thin ice* were shown it without any context, why should they not take it literally? In fact, in contextualised uses, both the literal meaning and the implications from context must operate together. In order to notice that a word string is incongruous *and* note that the meaning cannot be resolved, *and* register it as a fixed, or partly fixed expression, one must in some way interrogate it.

3.2 Idiom learning in a foreign or second language

One of the most persistent puzzles in applied linguistics is how learning another language can be so difficult, given that the same individual has already mastered one language (Wray, 2008b). A range of potential explanations has been extensively explored over the years. They include: a brief window (critical period) during which native-like proficiency can be achieved; one-off developmental calibrations that are anchored in the first language and do not change subsequently; differences in the extent and quality of exposure to another language; and differences in how the speaker is positioned emotionally and socially in relation to it (see Wray, 2002 for a discussion of these alternatives). The research findings are complex and often contradictory, not least because no two learners are quite the same.

That it becomes more difficult to master a second language as you get older is "not controversial" (Hakuta, Bialystok and Wiley, 2003: 31), nor that the most challenging aspect

for later learners is pronunciation (Flege, Yeni-Komshian and Liu, 1999). Kenneth Hyltenstam and Niclas Abrahamsson (2000) review a range of evidence from research studies, including many claiming to have found post-childhood learners with a fully nativelike level of mastery. Hyltenstam and Abrahamsson contend that although it is possible for highly proficient late learners consistently to be judged fully nativelike by native speakers, detailed linguistic analysis reveals differences not only in pronunciation but also lexical and grammatical production and intuitions. They concur with the 'fundamental difference hypothesis' that

> young starters acquire language through implicit, domain-specific mechanisms, while adult starters must use explicit, problem-solving strategies in L2 acquisition. (159)

The hypothesis further states that the latter learning approach is less effective overall.

The puzzle nevertheless persists: why do these explicit strategies impede the ultimate attainment of some adult starters more than others? For, notwithstanding subtle non-nativisms that are imperceptible without in-depth specialist analysis, there is a subset of post-childhood learners who do reach nativelikeness. Robert DeKeyser (2000) found that amongst Hungarian-speakers who immigrated to the United States after puberty, the level of nativelikeness in English correlated with the score on a verbal analytical ability test. This was not the case for those who immigrated at an earlier age. The implication is that "a high verbal aptitude [is] an asset, for example, in conscious reflection on grammatical rules" (Hyltenstam and Abrahamsson 2000: 159). DeKeyser goes on to argue that it is therefore vital to provide post-childhood learners with form-focussed input. Not to do so is "[to] deny learners with high analytic ability the use of the only mechanism at their disposal to master certain basic structures in the L2" (520).

This being so, we can infer that the major impediment to successful nativelike mastery of an adult-learned language would be the absence of this "high analytic ability". The argument could thus be made that, once the critical period for implicit learning has passed, it would be most cost-effective to screen potential language learners for the capacity to learn explicitly.[3] Such an extreme position, however, could underestimate the availability of implicit learning approaches in immersion contexts, where other factors, such as integration and identity, may play the dominant role in how well the learner accommodates to the speech patterns of surrounding native speakers.

The risk, when taking the explicit, analytic approach to learning, is the adoption of an over-regulated, piecemeal approach that seeks a greater measure of regularity and logicality than natural languages actually display. Alison Wray and George Grace (2007) review and discuss evidence that languages spoken only by native speakers accumulate irregularities (see also Trudgill, 2011; Wray, 2008a), while, conversely, languages become more regular if they are used extensively by post-childhood second language learners. This dynamic underscores the potential for a language that is under the social control of native speakers to repel adult learners by undermining their efforts to find regularity and predictability, while a language that is under the social control of the incomers is progressively divested of irregularities, making it easier for the next wave of incomers to learn. It is consistent with this view that, in lingua franca contexts, expressions with a non-literal meaning have a disproportionate potential to derail communication (e.g. Seidlhofer, 2011; Mackenzie, 2014).

[3] Wray (2008b) also considers the value of delaying language teaching in schools until all class members have exited the implicit learning period, to avoid teaching learners with fundamentally different approaches to learning.

4 Investigating how unknown idioms are interpreted

Using a think-aloud method, Hongshan Zuo (2008) investigated how Chinese learners of English of different proficiency levels tackled the comprehension of unknown idioms. She identified ten strategies. By far the most prevalent, at 62.5% of all attempts at meaning-making, was what she terms 'making schematic inferences' – that is, drawing on the contextual information in the passage containing the target item. Zuo found that the participants with higher English proficiency used more context than those with lower proficiency. The investigation below is in some ways similar to Zuo's. However, the aim was to compare the strategies of native and non-native speakers.

4.1 Selecting suitable idioms expressions

Three possible approaches could be taken to investigate whether first and post-childhood second language users decode unknown idioms in the same way. One is to compare native speaker children with non-native speaker adults. Here, one could anticipate that expressions known to native speaker adults might not yet be known by either set of participants. However, as we saw earlier, comparing children with adults would introduce many unwanted variables, including cognitive maturity and knowledge of the world. A second option is to give native and non-native speaker adults stimuli from an artificial language. But this is ecologically far removed from real, functional language learning. The third possibility is to find stimuli with formulaic characteristics that have not been encountered by adult native or non-native speakers before, but that native speakers will nevertheless accept as legitimate in their language.[4]

Such items can be found in the work of the British author Georgette Heyer (1902-1974), thirty-four of whose novels are set in the Georgian and Regency period. Heyer made a particular study of the turns of phrase used at the time, and incorporated examples into the dialogue of her characters (Hodge, 1984: 39-40). The expressions are semantically opaque, usually because they are metaphorical, and thus they signal themselves – to adults at least – as formulaic. Yet since Heyer does not explain them, the reader has to infer their meaning somehow. Consequently, a native speaker who reads Heyer's Georgian and Regency novels undergoes a similar experience to that of a proficient non-native speaker reading works in the second language in which there are formulaic expressions that are unfamiliar. Example 2 gives a flavour of Heyer's text, from *The Toll Gate* (1954).

Ex.2:

Chirk dived a hand into his pocket, and drew forth a snuff-box. It was a handsome piece, as its present owner acknowledged, as he offered it, open, to John.

'Took it off of a fat old gager a couple o' years back,' he explained, with engaging frankness. 'Prigged his tattler, too, but I sold that. I'm a great one for a pinch o' merry-go-up, and this little box just happened to take my fancy, and I've kept it. I daresay I'd get a double finnup for it, too,' he added, sighing over his own prodigality. 'It's worth more, but when it comes to tipping over the dibs there ain't a lock as isn't a hob-grubber.'

[4] If a phrase is unfamiliar to a native speaker, it fails one of the tests normally applied to establish formulaicity (Wray and Namba, 2003; Wray, 2008a): having a known conventionalised meaning. But other tests will confirm it as formulaic.

Depending on whether expressions are counted as a single item or several words, and which words are considered to be opaque (for example, are 'pinch', 'double' and 'lock' to be taken literally or not?– see later discussion), between 4% and 6% of the text will be unknown to most native speakers. According to Batia Laufer and Geke Ravenhorst-Kalovski (2010) a reader must understand between 95% and 98% successfully to understand a text.

4.2 Participants

Twenty-three non-native speaker[5] participants were recruited. They were overseas undergraduates, postgraduates and academic visitors at two British universities. Their proficiency on arrival in the UK had varied from around IELTS 6.0 to near-native, and ten first languages were represented. In the absence of an up-to-date measure of their proficiency that would take into account their learning since arriving in the UK, an informal task was compiled to examine, specifically, their familiarity with some common English idioms (see Appendix 1).[6] The non-native speakers' scores spread fairly equally from 0 to 14 (see Figure 1). The mean was 8.48, the median 9.

Seventeen native speakers – eleven undergraduates and six teachers of English as a foreign language at the same two universities – also participated. They also took the English idiom completion test, and all of them scored 15 out of 15.

–Figure 1 about here—

4.3 Design

From a longer list assembled from Heyer's novels *The Tollgate, The Black Moth* and *The Reluctant Widow*, we selected ten expressions[7] that displayed features likely to be of interest (Table 1). Most of the expressions were figurative, but two were not. Rather, they were literal, but used words that the participants would need to understand in order to realise that fact: compare *prigged his tattler* – stole his watch (Partridge, 1973) in Example 1.

–Table 1 about here–

The expressions, written in bold, were typed onto paper, in their original paragraph, lightly edited to make the text freestanding (see Appendix 2). Participants were interviewed individually in a quiet room, and were audio-recorded as they read the passages aloud and attempted to explain the expressions. The participant saw the items in one of two different orders, so as to reduce the impact of fatigue, practice or strategy-development on the combined results.

5 Results

The transcribed material was coded for nine features, which are listed and illustrated from

[5] Our choice of the term 'non-native speaker' is deliberate, even though in some contexts it underrates the legitimacy of the second language speaker as an expert user. We are interested in whether the fact of learning a language after infancy and in a classroom impacts on how unknown expressions are interrogated.
[6] This task has no formal status in language testing and we cannot by any means take it that increased performance in this task would exactly mirror a wider measure of proficiency. However, we will treat their test performance as a proxy for proficiency for the purposes of this account.
[7] The meanings of the expressions are given in Appendix 3.

the data in Table 2.

--Table 2 about here--

In the analysis, an individual could be assigned any number of strategies. However, reuse of the same strategy for one expression was only counted once. That is, if a participant gave two different analogies based on the same phrase, the A category was given a score of 1, not 2 for that item. Thus, when the strategy choices were combined for each group, the maximum possible score for each strategy was 17 (one each by the native speakers) and 23 (one each by the non-native speakers). To make these values compatible for calculations and graphs, the scores were turned into percentages (Table 3). It should be kept in mind that percentages can be misleading when the underlying figures are small. For example, a single response by one native speaker counts as 5.89% whereas a single response from a non-native speaker is 4.35%.

--Table 3 about here--

Many of the values are small – that is, a strategy was used occasionally, by one or two people. As it is important not to read too much into such occurrences, the discussion below will focus on those strategies that figured most prominently: drawing on context (C), identifying analogies (A), and focussing on unknown items within the lexical target (LT). The category L (links to lexis outside the target) involves several sub-types of response, and will not be discussed here for lack of space.

5.1 The role of context

--Figure 2 around here--

Just as in Zuo's (2008) study, for both groups, the context was the main basis for working out the meaning. Figure 2 shows that the non-native speakers referred to the context less than the native speakers in all but two cases (*at home to a peg* and *cut a wheedle*). The difference in context use between the groups was statistically significant ($t = 2.8905$, $df = 9$, $p = 0.0179$, two-tailed).

In order to examine why some items were less reliant on context than others, the ten stimulus expressions were grouped in two ways, according to: (a) the amount of helpful context they had in the surrounding text; (b) the familiarity of the words within them. Examples 3 to 5 illustrate how participants drew on contextual cues in high, medium and low context/transparency environments.

Ex.3
HIGH CONTEXT, MEDIUM TRANSPARENCY: I think *under the hatches* must be an expression to mean *being under hand,* to be *working behind the scenes* and I think that's because where it says *your cousin was a common thief,* thieves obviously won't do it in plain sight they'll try and do it covertly (NS ML102).

Ex.4
MEDIUM CONTEXT, LOW TRANSPARENCY: *Not to cut a wheedle* with you. I'd say I'd just be thinking that's like well *I don't want to fall out with you over this*. [The] use of the word *cut* is quite an aggressive word and then *not* and you show that they don't want to seem like they're arguing and then where *I put it more strongly than that* showing that they

do that there's some minor disagreement. It's kind of like they're hedging first (NS FL112).

Ex.5
LOW CONTEXT, HIGH TRANSPARENCY: I think [*handle the ribbons*] means an ability to handle some difficulties because the word *handle* makes me feel like oh it should be something that hard to deal with and I think in the previous sentence like *greater importance* so I assume that it should be something really important and probably not something easy to do (NS FL204).

Figures 3 and 4 show that there was no particular tendency for passages with greater or lesser contextual information, and greater or lesser transparency, to affect how much the context was relied on.

--Figures 3 and 4 around here--

5.2 Analogy

The native speakers were more likely than the non-native speakers to come up with analogies as a way of determining a meaning. Examples 6 to 9 show that analogies were typically in the form of another idiomatic expression.

Ex.6
I think *never had a feather to fly with* must mean the same thing as *didn't have two coins to rub together* (NS ML102)

Ex.7
Handle the ribbons it sounds very much like *to know the ropes* so I assume it means the same thing (NNS FL208)

Ex.8
[*At home to a peg*] it's funny, sometimes other idioms come to mind so *to a tee* whereas it's nothing really to do with that I don't think. It's just words, phrases of a similar structure come to mind (NS FL118)

Ex.9
So *under the hatches* so like *under the wing* maybe or *over-protected* in this context because or like um yeah over-protected so he never really got into contact with any like crime or whatever so he doesn't steal *under the hatches* maybe he was his parents or whatever not letting him take part in the you know youth culture or whatever it is (NNS ML209)

Figure 5 compares the use of analogy by the two groups. The difference was statistically significant ($t = 3.0393$, $df = 9$, $p = 0.014$).

--Figure 5 around here—

One reason for the greater use of analogy by native speakers might be their greater familiarity with alternative ways of expressing an idea. If so, then we would expect to find an increase in the use of analogy in the non-native speakers according to increasing proficiency.

A Spearman's rank correlation test was used to establish whether there was a relationship between proficiency and the amount of analogy used. The result was highly significant (R = 0.66538, p = 0.00053). This finding suggests that as non-native speakers come closer to native-like competency, they increasingly draw on analogy to help them work out the meaning of an unknown expression. Analogy has long been cited as a major means by which effective language learning takes place (e.g. Rumelhart and McClelland, 1986).[8]

5.3 Admitting to unknown words

It might seem obvious that, when trying to work out the meaning of an expression, one would mention any individual words that one either was not familiar with, or that one did not believe could have their usual meaning (Examples 10 to 12).

Ex.10
I don't know what *brown* could mean but I'm guessing it means you're being *too forward* (NS FL103)

Ex.11
Wheedle, what is that? Is it something to sew? (NNS FL207)

Ex.12
Culp, I have no idea what that means. Maybe it's *load* or something (NNS ML209)

Although both groups did refer to unknown words, there was a striking difference in how often. Of the 170 possible opportunities for the 17 native speakers to respond in this way across the ten stimuli, only four featured this strategy (2.35%). For the 23 non-native speakers, there were 28/230 instances (12.17%). Of course, not all of the expressions contained a word that was unknown. Having said that, none of the expressions was actually transparent, which means that there was potential for the meaning of any word to be questioned. In the final section, we discuss in more depth what might lie at the heart of the differences in the approaches to decoding taken by the native and non-native speakers.

6 Discussion

A fundamental feature of humans' language ability is the capacity to understand sentences never seen before. By definition, a new word string is not familiar or formulaic and this means that we do not have any opportunity to look it up as a single lexical entry and find a meaning for it as a whole.[9] Instead, we need to work out its meaning, and, as noted earlier, we do this by bringing together two types of information: internal semantics derived from the meanings of the individual parts and their grammatical relationship, and externally determined meaning derived from context.

[8] However, when Zuo (2008) compared the strategies used on unknown English idioms with more and less direct equivalents in Chinese, she found no major differences, suggesting that analogy was not playing a major part for her participants.

[9] Technically this is not quite true, because a formulaic frame could be populated with novel slot-fillers, e.g. *The Emperor penguin pulled the snail's leg.* A truly novel sentence is one that does not deploy a formulaic frame, e.g. *Leave seven potatoes and sign the undertaking.* Having said that, emergent grammar models hold that any grammatical sentence is based on a frame. But this need not concern us here, because our focus is on the semantics rather than the grammar.

Where the two agree – that is, the composite meaning of the word string fits the context – we will take the meaning to be literal. Where there is a conflict, we will usually believe the context and look for an alternative, figurative meaning. Thus, the expression *have no feather to fly with* cannot be taken literally in the context of describing a human – for even though it is literally true, it is not a relevant observation to make, pragmatically. Consequently, the reader will infer that it is a figurative expression, and the interpretation assigned will be based on finding a parallel meaning (Example 13).

Ex.13

I'd say that's like a big exaggeration trying to say that he's *never even had a feather to fly with* so like why would he have any money to do anything when he never even had something of little value like a feather that's going to help him along (NS FL112).

When participants could make no sense of the expression either literally or figuratively because it contained an unknown word or because the words, though recognisable, could not be assigned a relevant meaning, they ranged wide in their attempts to come up with an interpretation. Where words had more than one potential meaning, they struggled to know which one to favour. For example, several respondents were unsure whether to read *yard* as a measurement or as a place, and *bustle* as part of a woman's clothing or as 'busyness'. What might determine when a word was trusted to have its typical meaning (albeit then used figuratively) and when it was not? Why was a participant likely to decide that *cut* did mean *cut* but that *wafer*, *bustle* and *brown* might not mean what they appeared to? If we consider the full set of words that variously caused difficulty for the participants (*hatches; peg; culp; wafer; brown; wheedle; yard; tin; ribbons; bustle*) we find that they are all nouns and adjectives except *culp*, which no one would have known. Of the words that might have caused problems, but that were taken at face value (*under; at home; feather; fly; do; cut; handle; cry; rope; drawing*), two are nouns, one is a preposition, one is a prepositional phrase; the remaining six are verbs. So it may be that when we encounter an unknown expression, we tend to place more faith in the meaning of verbs than nouns. Certainly, it is recognised that not all parts of speech are the same in terms of how we process them (Wray, 2015: 749).

Finally, we return to the finding that native speakers were less likely to mention words they didn't know the meaning of. What does this signify? Several explanations are possible. One is that they encountered fewer unknown words in the first place. However, this cannot be the whole explanation. Focussing on the two examples containing words that would have been unknown to all participants, *culp a wafer* and *cut a wheedle*, the difference persists. Although there were 34 opportunities for native speakers to mention not knowing one of the words, the admission was made only twice (5.88%). In contrast, in the non-native group the incidence was 14/46 (30.43%).

A second possible explanation relates to confidence, though with two possible consequences. One is that the native speakers had more confidence than the non-native speakers that words unknown to them (including archaisms and localisms) were genuinely difficult to know. If so, they might have taken it as read that they didn't need to tell the interviewers (native speakers) that they didn't know the word, because the obscurity was shared knowledge. The other possibility is that the native speakers *lacked* the confidence to admit not knowing the meaning of a word in their own language, whereas the non-native speakers were more accustomed to saying that they did not know an item.

The third potential explanation is that the lower incidence of mentioning unknown words is linked to the native speakers' greater use of context. In a study of reading habits, Jan Hulstijn (1993) found that second language learners looked up fewer unknown words if the

context provided them with sufficient information to allow them to proceed without glossing them. Context could assist in two ways here. First, it could generate a springboard for inferring sufficient meaning to manage – such as knowing that the unknown word refers to a tool, type of food, or whatever, and not seeing it as necessary to know more than that.[10] Second, the (relative) completeness of the sentence meaning without the unknown item could signal that its inclusion was primarily for effect rather than content. In such cases the reader would only need to know that it conveys, say, annoyance or doubt. These 'make-do' strategies represent a version of what Wray (2002) terms 'needs only analysis' – only break down material if you have to, and only as far as you need do.

What we may be seeing here is that native speakers had greater trust in their interpretation of the context to patch up any lack of knowledge about words, while the second language learners were likely to want the word to elucidate their interpretation of the text as a whole. If so, then we are brought full circle, back to the research literature on formulaic language. It was noted earlier that in our first language we may deal in larger units, and that second language learning, particularly after childhood and/or in a classroom, encourages a greater focus on the internal make up of word strings. If so, then it would follow that the non-native speakers, perplexed by an expression, would home in on the word they didn't know, and talk about it. The native speakers, on the other hand, may simply not see any significance in not knowing the *word* when they have already realised that they don't know the *phrase* in its context.

In sum, there are some indications that it is a mark of one's knowledge of, and confidence in, a language, that one can transcend individual items within an unknown phrase, to deduce meaning from the context and through analogy. Meanwhile, the fact that the non-native speakers increasingly approximated this pattern according to their greater familiarity with the target language – at least as measured by the capacity to complete common idiomatic expressions – suggests that, in this particular regard at least, post-childhood language learners do not remain different from native speakers in their approach to language processing, when they command the second language to a very high level.

References

Dąbrowska, Ewa and Lieven, Elena (2005). 'Towards a Lexically Specific Grammar of Children's Question Constructions.' *Cognitive Linguistics* 16: 437-474.

DeKeyser, Robert M. (2000). 'The Robustness of Critical Period Effects in Second Language Acquisition.' *Studies in Second Language Acquisition* 22: 499-533.

Flege, James. E., Yeni-Komshian, Grace H. and Liu, Serena (1999). 'Age Constraints on Second Language Acquisition.' *Journal of Memory and Languag*e 41.1: 78-104.

Grice, H. Paul (1975). 'Logic and Conversation.' *Syntax and Semantics, Vol. 3.* Eds P. Cole and J.L. Morgan. New York: Academic Press. 41-58.

Hakuta, Kenji, Bialystok, Ellen and Wiley, Edward (2003). **'**A Test of the Critical-period Hypothesis for Second-language Acquisition.' *Psychological Science* 14.1: 31-38.

Hodge, Jane Aike (1984). *The Private World of Georgette Heyer*. London: Bodley Head.

[10] The capacity to infer the meanings of unknown words in this way has been demonstrated in dogs. Kaminski, Call and Fischer (2004) report that a border collie, Rico, "knew the labels of over 200 different items. He inferred the names of novel items by exclusion learning and correctly retrieved those items right away as well as 4 weeks after the initial exposure" (1682).

Hulstijn, Jan H. (1993). 'When Do Foreign-language Readers Look Up the Meaning of Unfamiliar Words? The Influence of Task and Learner Variables.' *Modern Language Journal*, 77.2: 139-147.
Hyltenstam, Kenneth and Abrahamsson, Niclas (2000). 'Who Can Become Native-like in a Second Language? All, Some, or None?: On the Maturational Constraints Controversy in Second Language Acquisition.' *Studia Linguistica* 54.2: 150-166.
Kaminski, Juliane, Call, Josep and Fischer, Julia (2004). 'Word Learning in a Domestic Dog: Evidence for "Fast Mapping."' *Science* 304: 1682-3.
Laval, Virginie and Bernicot, Josie (2002). '"*Tu es dans la lune":* Understanding Idioms in French-speaking Children and Adults.' *Pragmatics* 12.4: 399-413.
Laufer, Batia and Ravenhorst-Kalovski, Geke C. (2010). 'Lexical Threshold Revisited: Lexical Text Coverage, Learners' Vocabulary Size and Reading Comprehension.' *Reading in a Foreign Language* 22: 15-30.
Lieven, Elena, Behrens, Heike, Speares, Jennifer and Tomasello, Michael (2003). 'Early Syntactic Creativity: A Usage-based Approach.' *Journal of Child Language* 30: 333-370.
MacKenzie, Ian (2014). *English as a Lingua Franca: Theorizing and Teaching English*. London: Routledge.
Moon, Rosamund (1998). *Fixed Expressions and Idioms in English: a Corpus-based Approach*. Oxford: Clarendon Press.
Nippold, Marilyn A. (2006). 'Language Development in School-age Children, Adolescents and Adults.' *Encyclopedia of Language and Linguistics.* Ed. K. Brown. Oxford: Elsevier. 368-373.
Partridge, Eric (1973). *Routledge Dictionary of Historical Slang,* 6th ed. London: Routledge.
Peters, Ann M. (1983). *Units of Language Acquisition.* Cambridge: Cambridge UP.
Rumelhart, David E. and McClelland, James L. (1986). 'On Learning the Past Tenses of English Verbs.' *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 2: Psychological and Biological Models.* Eds D.E. Rumelhart and J.L. McClelland. Cambridge, MA: MIT Press. 216-271.
Seidlhofer, Barbara (2011). *Understanding English as a Lingua Franca*. Oxford: Oxford UP.
Taylor, John R. (2012). *The Mental Corpus*. Oxford: Oxford UP.
Tomasello, Michael (2003). *Constructing a Language: A Usage-based Theory of Language*. Cambridge, MA & London: Harvard UP.
Trudgill, Peter (2011). *Sociolinguistic Typology*: *Social Determinants of Language Complexity*. Oxford: Oxford UP.
Vulchanova, Mila, Vulchanov, Valentin and Stankova, Margarita (2011). 'Idiom Comprehension in the First Language: A Developmental Study.' *Vigo International Journal of Applied Linguistics* 8: 207-234.
Wray, Alison (2002). *Formulaic Language and the Lexicon.* Cambridge: Cambridge UP.
Wray, Alison (2008a). *Formulaic Language: Pushing the Boundaries*. Oxford: Oxford UP.
Wray, Alison (2008b). 'The Puzzle of Language Learning: From Child's Play to "Linguaphobia."' *Language Teaching* 41: 255-273.
Wray, Alison (2015). 'Why Are We So Sure We Know What a Word Is?' *Oxford Handbook of the Word.* Ed. J. R. Taylor. Oxford: Oxford UP. 726-750.
Wray, Alison and Grace, George W. (2007). 'The Consequences of Talking to Strangers: Evolutionary Corollaries of Socio-cultural Influences on Linguistic Form.' *Lingua* 117: 543-578.
Wray, Alison and Namba, Kazuhiko (2003). 'Formulaic Language in a Japanese-English Bilingual Child: A Practical Approach to Data Analysis.' *Japanese Journal for Multilingualism and Multiculturalism* 9: 24-51.

Zuo, Hongshan (2008). 'Comprehension of Unfamiliar English Idioms by Chinese EFL Learners in Reading.' *Journal of the Chinese English Language Association* 31.3: 9-29, http://www.celea.org.cn/teic/79/080801_9.pdf

Appendix 1: English idiom familiarity test

Name: **Age:** **First language:**
Other L2s:

Use the two words in the middle to write a phrase that fills the gap in the sentence. You can change the words, or add more words, to make a complete phrase.

Example:

I think I [MAKE, MISTAKE] in my exam yesterday.	made a mistake

1. I'm not sure, but [FAR, KNOW] Heathrow is the busiest airport in Europe.	
2. Jane is 10 years old this year. [SEEM, YESTERDAY] that she was a baby.	
3. Listen to Martin reading! [HARD, BELIEVE] he is only five years old.	
4. They tried everything, but [NO, AVAIL].	
5. It was such a surprise. At first I couldn't [TAKE, BOARD]	
6. At first I didn't recognise him, but [TURN, OUT] we went to the same college years ago.	
7. Anyway, [STORY, SHORT] we missed the bus.	
8. [FOR, KNOW] it could be something completely different.	
9. [CHANCE, ARE] it won't happen.	
10. Call the police. You can't [LAW, HANDS]	
11. She started at the bottom of the company, but [WORK, WAY]	
12. She's only a young child. That book [OVER, HEAD]	
13. A: You're singing that song again. B: I know. I can't [GET, HEAD]	
14. [LINE, WORK] are you in, John?	
15. It's [LITTLE, CONCERN] to them.	

Appendix 2: Stimuli

1

'What could there be that anyone should want?'
'I don't know, but I'll swear there is something. Of course, it may not be a paper; I wonder if Eustace had stolen something of value? He was always **under the hatches**, and–'
'I will not allow it to be possible!' said Elinor. 'Do you wish me to believe that your cousin was a common thief?'

2

Peter was **at home to a peg**, for this was his village of boyhood memory. He recalled days on the wing spent in the company of his late grand-uncle from that sunny place, and his own father who, carting his load of turnips to the city markets, had often brought him with him, seated in a special hollow made amongst the turnips. The joy was his when they stayed overnight at Mrs Molloy's lodging house in the haymarket.

3

'Are you imagining that you have become a rich woman overnight?' Carlyon enquired. 'I wish it may be so, but I fear it will be no such thing. You are more likely to discover that you are liable for God knows how many debts.'
'Lord, yes!' said Nicky cheerfully. 'Eustace never **had a feather to fly with**!'

4

'You enjoy shooting then?' Freddy was saying as he sipped his wine.
'For the pot, yes,' Dick agreed easily. 'Can't see the point in it otherwise. Waste of powder and shot.'
'Are you a good shot then, sir?' Rookwood asked, leaning forward slightly to look down the table at him.
'Passable,' Turpin replied vaguely.
'Come, Richard, you can **culp a wafer** with the best of them,' Glenrae put in deliberately.

5

Her eyes began to dance, and her lips to quiver. 'You know, you are the most provoking creature I ever encountered!' she told him.
'Oh, come, now, that's **doing it rather too brown**!' he expostulated. 'Remember, I was acquainted with your brother Rowland!'

6

'It's very good of you to say so, Duke!' responded Sir Nugent, acknowledging the tribute with a slight bow. 'I don't mind admitting it wasn't easy. It took a lot of effort. If there is a thing I pride myself on it's that. 'Lady Henry,' I said – well, not **to cut a wheedle with you**, Duke, I put it more strongly than that! 'My love,' I said, 'we shan't make him happy if we keep His Lordship waiting at the rendezvous. Take my word for it!' She did.'

7

In the years around 1776, there was a tollgate at Botley Hill, at a place called Coldharbour Beeches. The road was lined with beech trees on each side, and ran along the boundary of the nearby parish of Tatsfield, near the borders of Kent and Surrey. The toll gate cottage is still standing, and still occupied, but the toll gate, which could have blocked the road, has long since gone. The name of the cottage is now Paygate cottage. Stagecoaches paid toll, but mail

coaches did not. At a blast on **the yard of tin**, the toll gatekeeper would hurl himself out to open the gate, as the coach thundered through. Every type of traffic gave way to the mail coach.

8

There is now no branch of a lad's outdoor education of greater importance, not only to himself, but also to others, than that of an ability **to handle the ribbons**.

9

'All right, be bloodthirsty if you must. Boys will be boys. That leaves only the question – who and how do we handle disposing of the wretched girl?'

'That's two questions. I don't know how to do it, but I do know who. I've thought this out most carefully. We both do the deed. That way neither of us is apt **to cry rope on the other**.'

There was a short silence while his co-conspirator weighed his latest suggestion.

10

When the cloth had been removed, and the brandy placed on the table, Stacy said, with his air of rueful frankness: 'I must tell you, sir, that I was devilish glad to get your letter! I've **been drawing the bustle** a trifle too freely.'

Appendix 3: Meanings of stimulus phrases

(sources: http://www.georgette-heyer.com/slang.html;
http://candicehern.com/regency-world/glossary/
http://www.regencyassemblypress.com/Regency_Lexicon.html

- *Under the hatches* – in debt
- *at home to a peg* – very at home
- *have no feather to fly with* – have no money
- *culp a wafer* – hit a small object (wafer)
- *do it too brown* – overdo it, so it's not credible
- *cut a wheedle* – ingratiate self with someone by lying
- *yard of tin* – post horn
- *handle the ribbons* – drive a coach or carriage
- *cry rope on* – give [someone] away; tell secret
- *draw the bustle too freely* – spend too much money

Table 1: Expressions used in the study

Expression	Figurative	Unknown word in L1
under the hatches	✓	X
at home to a peg	✓	X
[not have a] feather to fly with	✓	X
culp a wafer	X	✓
do[ing] it rather too brown	✓	X
[not] to cut a wheedle with you	✓	✓
the yard of tin	X	X
to handle the ribbons	✓	X
to cry rope on	✓	X
draw[ing] the bustle [a trifle] too freely	✓	(✓)

Table 2: Coding of responses (NS = native speaker; NNS = non-native speaker)

C	Use of **context** to work out the meaning, e.g. 'they're on about shooting …maybe *culp a wafer* is shooting a good shot' (NS FL103)
LT	Mentioning that one of the **lexical items in the target expression** was not known, or evidently had a different meaning from a known one, e.g. 'I don't know what a peg is' (NNS EJF)
A	Pinning down meaning by means of an **analogy**, e.g. '*doing it rather too brown* is *that's rich coming from you*' (NS ML101); see also the example under P.
L	Making conceptual links using **lexis** not in the target, e.g. [*feather to fly with*] 'Maybe birds or something like that' (NNS FL209)
P	Making a **phonological** link with one of the words in the target expression, e.g. '*under the hatches* … I know in modern day we say *bury the hatchet*' (NS FL103)
M	Creating or extending a metaphor, e.g. '*take the reins, handle the ribbons*, yeah, importance of leadership' (NS FL111)
GP	Commenting on **grammar, semantics, pragmatics** or some other mechanism for working it out, e.g. 'So that 'for' explains the reason why he was *at home to a peg*' (NNS SSB)
DK	Giving up, e.g. 'I'm not quite sure about this phrase' (NNS FL221)
TR	**Translation** from another language, e.g. 'in terms of my first language German maybe *wafer* could be something like weapon, *Waffe*' (NNS ML209)

Table 3: Occurrence of each strategy type (percentages)

		C	LT	A	L	P	M	GP	DK	TR
Culp a wafer	Native	100	11.76	5.89	0	5.89	0	5.89	0	0
	Non-native	91.3	56.52	13.04	0	0	0	8.7	8.7	0
Under the hatches	Native	88.24	5.89	52.94	29.41	0	0	0	0	0
	Non-native	76.26	47.82	17.39	8.7	4.35	0	4.35	4.35	4.35
At home to a peg	Native	82.35	35.29	52.94	47.06	0	0	0	5.89	0
	Non-native	86.96	60.87	13.04	17.39	4.35	0	4.35	8.7	0
Feather to fly with	Native	94.12	41.18	47.06	0	0	11.76	0	0	0
	Non-native	86.96	52.17	4.35	4.35	0	4.35	4.35	0	4.35
Do it too brown	Native	88.24	41.18	5.89	5.89	11.76	0	0	0	0
	Non-native	82.61	52.17	4.35	0	4.35	0	0	8.7	8.7
Cut a wheedle	Native	76.47	17.65	41.18	0	0	0	0	5.89	0
	Non-native	82.61	39.13	17.39	0	4.35	0	4.35	4.35	4.35
Yard of tin	Native	100	64.71	11.76	11.76	0	0	0	0	0
	Non-native	86.96	73.91	0	0	0	4.35	0	0	0
Handle the ribbons	Native	70.59	82.35	35.29	0	5.89	11.76	0	0	0
	Non-native	65.22	78.26	26.08	0	0	8.7	4.35	0	0
Cry rope on	Native	94.12	35.29	17.65	5.89	0	0	0	0	0
	Non-native	86.96	91.3	17.39	0	0	4.35	0	0	0
Drawing the bustle	Native	100	41.18	11.76	5.89	5.89	0	0	0	0
	Non-native	86.96	39.13	0	0	8.7	0	4.35	8.7	0

List of figures

Figure 1: Non-native speaker scores on the English idiom familiarity test

Figure 2: Use of context to interpret expressions

Figure 3: Use of context, grouped by amount of context available

Figure 4: Use of context, grouped by transparency of lexis

Figure 5: Use of analogy to work out meaning