Applied Linguistics Advance Access published March 12, 2011
Applied Linguistics 2011: 1–23
doi:10.1093/applin/amr009

Difficulties in Metaphor Comprehension Faced by International Students whose First Language is not English

*JEANNETTE LITTLEMORE, PHYLLIS TRAUTMAN CHEN, ALMUT KOESTER, and JOHN BARNDEN

University of Birmingham
*E-mail: j.m.littlemore@bham.ac.uk

This article reports a study on metaphor comprehension by the international students whose first language is not English, while attending undergraduate lectures at a British university. Study participants identified words or multiword items that they found difficult in extracts from four academic lectures, and they interpreted metaphors from those extracts. Among the items reported as difficult, we established the proportion of metaphorical items, plus the proportion of items composed only of words familiar to the students. We developed a measure of the extent of students' awareness of their metaphor interpretation difficulties, plus a scheme for categorizing the most common types of metaphor misinterpretations. We found that, of the items that were difficult though composed of familiar words, ~40 per cent involved metaphor. Further, when the students misinterpreted metaphors, they only seemed aware of having difficulty in ~4 per cent of cases. As university lecturers use metaphors for important functions, such as explaining and evaluating, such international students may thus be missing valuable learning opportunities. Our error categorization scheme could be used in helping English learners with metaphor comprehension.

INTRODUCTION

In a previous study (Littlemore 2001), it was found that international students whose first language is not English (henceforth 'international students') often experience difficulties understanding metaphor used in university lectures. This is problematic, because metaphor assists with important pedagogical functions such as description, explanation, and evaluation (Corts and Pollio 1999; Cameron 2003). Difficulties in understanding metaphor, therefore, can hinder a student's ability to follow the academic content of a lecture, as well as to grasp the lecturer's stance towards the material presented.

Littlemore's study revealed two types of metaphor comprehension difficulties: *misunderstanding* and *non-understanding*. With misunderstanding, people believe they have understood correctly and therefore do not seek clarification. In the study, which focused on 10 evaluative metaphors used in a lecture

attended by 20 international students, most of the students misinterpreted at least one of the metaphors, often seriously affecting their understanding of the lecturer's position. Additionally, metaphorical usages throughout the lecture accounted for most of the items (145 out of 180) that they found difficult to understand.

The current study builds directly on Littlemore's earlier one. Its purpose is to measure more fully the extent to which, and in what ways, metaphor presents problems to international students in a British university setting. The original study was narrow in its selection of participants (all were Bengali-speaking post-graduate civil servants attending an International Development course). We wanted to see if similar results would be obtained with students at an earlier academic level from a range of linguistic backgrounds. For the current study, we chose students enrolled in a British university pre-undergraduate International Foundation programme in Arts and Social Sciences; and we investigated their metaphor comprehension in four lectures. The aim of the International Foundation programme is to prepare international students whose first language is not English for undergraduate study in the UK. It is a 1-year programme for academically able students whose school leaving qualifications are not recognized for direct entry to degree programmes in the UK. The overall aim of the programme is to help the students to develop their knowledge of relevant subject matter so that they will be able to engage in undergraduate study the following year.

As in the earlier study, we asked the participants to identify any words or word clusters that they found difficult to understand in some lectures. We wanted not only to determine the proportion of metaphorical items among all items reported as difficult, but also to establish whether the problems posed by these metaphorical items were due to their metaphoricity, as opposed to the students simply being unfamiliar with words within them.

Another aim of our research was to explore the combined issues of metaphor non- and misunderstanding. We asked the participants to explain in writing metaphors from the lectures that we had selected. We then developed an error categorization scheme for use in analysing their metaphor explanations. This scheme identified eight different types of metaphor interpretation errors and enabled us to categorize any given interpretation as acceptable or not, according to either strict or generous criteria. We were then able to use our analyses as the basis for quantifying the extent of metaphor understanding (or lack thereof) for each lecture.

The final aim of our research was to quantify the extent of metaphor misunderstanding. We compared the metaphorical items that students had failed to interpret acceptably with the items they had initially reported as difficult to understand. The items that they had not reported, we reasoned, represented metaphor comprehension difficulties of which they were unaware. The percentage of such unreported difficult items out of the total of poorly interpreted metaphors served as our measure of metaphor misunderstanding in each lecture.

BACKGROUND

The importance of metaphor in educational discourse is well established (see, for instance, Littlemore 2001; Cameron 2003). Metaphor is suited to a wide range of teaching functions, among them description, explanation, exemplification, clarification, summation, restatement, and evaluation (Cameron 2003). In university lectures, evaluation is a particularly important use of metaphor (Drew and Holt 1998; McCarthy 1998; Littlemore 2001) and the evaluative component of a lecture is often cited as part of the rationale for this mode of teaching (Thompson 1994).

Additionally, metaphors can serve a meta-discursive function in lectures, sometimes referred to as 'agenda management' (Cameron 2003; Low *et al*. 2008). In this usage, they help signpost ongoing progress through the lecture, relating different parts of the lecture to each other (e.g. as main points, examples, asides, restatements, etc.). Examples from the data analysed in this study include: 'tomorrow we'll *wrap that up*' and 'the lecture is very much *built around* a number of slides'.

Such varied uses of metaphor suggest that it is neither a superficial nor a limited linguistic device in academic lectures. In a recent study of three social science lectures from the British Academic Spoken English (BASE) corpus, the metaphor density was found to be 10–13 per cent (Low *et al*. 2008). These relatively high figures suggest that the teaching of metaphor should be given importance in academic language preparation programmes.

Metaphor is notoriously difficult to define and identify (Steen 2007). In this study, we follow the Pragglejaz Group (2007) definition underlying the metaphor identification procedure (MIP) described there, in that we identify metaphor as any instance in which a particular lexical unit can be seen to have a more basic, contemporary sense in another context, and where the contextual sense can be understood in relation to that basic sense by means of comparison. For example, when one of the lecturers in our study says to the students: 'Last week we introduced the notion of the invisible hand', we consider his use of the word 'hand' to be metaphorical as he is clearly not talking about real, physical hands in this context; nevertheless the contextual meaning of the word can be understood in relation to the normal physical sense.

There are a number of different methods for manually identifying metaphor in text. We used two different methods in our study because of their relative suitability for different purposes. One of the methods was MIP (see above). Applying the MIP involves determining, first, the contextual meaning of a lexical unit, and then identifying its 'basic contemporary meaning'. If the former can be understood in relation to the latter by means of comparison, the lexical unit is judged to be metaphorical (e.g. *struggled*, in relation to political power). The criteria for basic contemporary meanings include the following: they refer to concrete entities or bodily actions, are precise rather than vague, and/or are historically older than the other meanings. For example, under these criteria, spatial prepositions, for example, *in*, would be

metaphorical if used in contexts other than physical space. Note that in MIP metaphoricity in any selected text is examined on an individual lexical-unit basis. Idioms, for example, are decomposed into separate lexical units, based on the rationale that the metaphoricity of an idiom, if it exists, is attached to specific words within it (e.g. *pop*, as in *pop the question*).

The other method we used was the vehicle identification procedure (VIP) of Cameron (2003). Unlike the MIP, which focuses on individual lexical units and is restricted to metaphor only, the VIP focuses, at least initially, on any 'stretch of language' whose surface meaning appears to be anomalous or incongruous with the surrounding co-text. The first step is to 'trawl' through a text for these language stretches, which could include idioms, metonymy, and other figurative language types. The next step is to test for metaphoricity by looking, first, for a 'domain incongruity' between the item and the topic to which it refers; then, secondly, for a way of resolving the incongruity in the context, such as through a meaning transfer. Depending on the specific research and discourse context studied, additional criteria might later be applied to separate out metaphorical items of interest and/or to decide on borderline cases—for example, technical language, animating metaphors, or metaphors related to grammatical form.

While these two approaches may result in slightly different metaphor density data and indices, either one produces a rough benchmark of the prevalence of metaphor in particular lectures. However, such benchmarks do not directly reveal the dimensions of the metaphor comprehension difficulties experienced by the students in our study. That is, a low-density count does not necessarily mean that metaphor will not be a problem for these students. Irrespective of the metaphor density, if students have difficulty in understanding the metaphors that help to convey the lecturer's key instructional or evaluative points, then the educational value of the lecture will be seriously diminished.

This problem is exactly what was found in the Littlemore (2001) study described above. Littlemore found that the misinterpretation of metaphors, arising from inappropriate connotations of the vehicle, sometimes led to strikingly different conclusions from what the lecturer intended. For example, a reference to the need for civil servants to *attack their jobs* was incorrectly understood as a need for them to take a critical stance toward their jobs rather than go about them enthusiastically. Even more problematically, in many cases the students were unaware that their interpretation was inadequate. They therefore did not seek clarification; if they had, their mistaken impressions might have been corrected. Compounding the problem of metaphor misunderstanding, research suggests that lecturers rarely explain the key metaphors used in their lectures to the students (Low *et al*. 2008).

Metaphor understanding is thus an important aspect of academic listening, contributing to a learner's comprehension of both the lecture content and the lecturer's stance. If international students whose first language is not English

do not understand the metaphors expressed in lectures, they not only miss out on these benefits, but risk leaving with erroneous concepts as well.

METHODOLOGY

Research questions

We organized our research according to four questions:

1 Of those language items that students found difficult to understand, what proportion were metaphorical?
2 Of those language items that students found difficult to understand despite being composed of familiar words, what proportion were metaphorical?
3 What kinds of errors did the students make, and how often did they make them, when asked to explain the contextual meaning of metaphorical items in the lectures?
4 To what extent were students aware of their errors in attempting to understand metaphorical items in the lectures?

Selection of participants and lectures

For our study, we chose a group of 20 students learning English on an International Foundation Programme at the University of Birmingham in 2007. Their countries of origin included China, Saudi Arabia, Kazakhstan, Cyprus, Japan, and Indonesia. All participants had an average IELTS score between 5.5 and 6.5 (with a minimum of 5.0 in all four skills). The majority of the participants had scores towards the lower end of this range. To develop their listening and note-taking skills, these students attended various undergraduate lectures in business, economics, language, and linguistics. These were regular lectures that were being offered to first year undergraduate students. The International Foundation Programme students sat in on the lectures as part of their programme but the lecturers were not asked to change their lecturing style in any way in response to the presence of these students. The number of International Foundation Programme students attending the lectures was very small in comparison with the large number of first year undergraduate students in the lectures, and as such is unlikely to have altered the way in which the lectures were delivered. Thus, the language used for analysis was fully contextualized language in use and not metaphorical language designed specifically for research. For the study, we selected the first four of these lectures that were taught by lecturers with English as their first language. Information about the lectures is shown in Table 1.

Table 1: Lectures used in the study

Lecturer and subject	Lecture length (min)	Transcript length (min)	No. of words in transcript	No. of students
Lecturer A (linguistics)	45:38	13:39	2114	18
Lecturer B (economics)	46:22	8:30	1454	4
Lecturer C (linguistics)	45:34	12:42	1914	12
Lecturer D (media)	51:19	11:56	1793	12

Student testing sessions

For each lecture, we ran a 1.5-h student testing session ~2 weeks after the lecture. Each session consisted of two main data collection activities, called Activity 1 and Activity 2. Both of these were based on a transcript provided to the students of one or more extracts from the lecture. Details concerning the length of the transcripts and the numbers of students attending the sessions are given in Table 1. In Activity 1, students listened to the extracts while marking up expressions which they had any sort of difficulty understanding. Then, in Activity 2, students wrote explanations of metaphorical items that we ourselves had previously identified in the transcripts. Students were not informed that the aims of the study were about metaphor. More details on both activities are given below.

One aim in selecting the extracts to be included in a transcript was to maintain uniformity across the lectures. Hence, each transcript started at the beginning of the lecture. The fact that the extracts studied came from the start of the lecture may have increased our chances of finding agenda management terms and attenuated the advantages and/or disadvantages for understanding that might come from listening to a final summarizing cluster of the sort found by Corts and Pollio (1999). This may have been a limitation of our study, and future studies could usefully focus on different lecture segments in order to access a different range of metaphorical functions. We also wanted to ensure that each extract was logically complete and could be understood during replay without any visual aids that might originally have been displayed. These various considerations explain why a single extract was possible from only one lecture (Lecture A), and why the Lecture B transcript was considerably shorter than the others. Time constraints did not permit the entire lectures to be used at the student testing sessions. The transcripts were presented in their original format, and no attempt was made to simplify them in any way for the benefit of the students.

Full student participation was not achieved in every session. After a turnout of 18 students at the first session, only four attended the next, due to a scheduling conflict. With the approaching end of the academic year, we conducted

the remaining two testing sessions back-to-back on the same day, with 12 students in attendance at each.

First data collection activity (Activity 1)

The first data collection activity at the testing session associated with each one of the four lectures, designated as Activity 1 for that lecture, addressed our first two research questions and part of the fourth. Students were given copies of the lecture transcript with a set of instructions (see online supplementary material for Appendix 1). They read through the transcript twice. During the first reading, they listened to the lecture videotape and, while doing so, underlined any words or word clusters that they found difficult in any way.

During the second reading, they reviewed their underlined items, using highlighter pens to indicate any words that were completely unfamiliar to them. This procedure enabled us later to eliminate any difficulties caused by unfamiliar words, as opposed to metaphorical uses of familiar words.

In our analysis of the data obtained, we scrutinized the metaphoricity of each word in the items the students had identified as difficult. For this process, we used the Pragglejaz MIP, with its focus on individual lexical units. MIP was preferable here to VIP because it is more readily applicable to short items that have already been picked out by other people. Using VIP would have required us first to make our own determination of metaphorical phrases in the overall text and then somehow to reconcile them with intersecting, student-chosen words or phrases.

Second data collection activity (Activity 2)

The second data collection activity in each testing session was designed to address research questions 3 and 4. Students were asked to explain in writing, using English or their own languages, the meanings of various metaphorical language items that we had chosen from the transcript used for Activity 1. Each item was presented to the students within a small amount of surrounding text from the transcript (see online supplementary material for Appendix 2 for instructions and examples). The students were not informed that the items were metaphorical. The average number of testing items selected for each lecture was 38. Using all of the possible metaphorical items in each transcript was infeasible, given the time constraints of the sessions.

The majority of the students chose to answer in English but there were 25 (out of a total of 1279) responses where the students provided answers in their own languages (Chinese, Russian, and Spanish). Translations were carried out in all these instances by members of the team in consultation with colleagues.[1]

In a first phase of metaphorical item selection for Activity 2, we followed VIP. That is, we trawled for short stretches of metaphorical language. The selection process involved an independent judgement by each researcher. Consensus was then reached through meetings and email exchanges.

The reason for using VIP rather than scrutinizing individual lexical units for metaphoricity as in MIP is that we hypothesized that students would find it easier to explain meaningful stretches of language than isolated lexical items. Also, phraseological meaning is important for lecture understanding, especially in the case of relatively fixed multiword expressions such as metaphorical idioms.

However, in a second phase, we narrowed the list of items by applying the MIP criteria to individual words within items. This helped us to clarify why we considered an item to be metaphorical, to achieve a good balance of conventional and unconventional metaphor, and to ensure that students would have a reasonable chance of recognizing the words and being familiar with their basic meanings. In unclear cases as to the basic meaning of an item, we consulted several dictionaries and discussed differences between them.

We gathered and kept in our final list a few items that were arguably cases of metonymy rather than metaphor. It is sometimes impossible to decide whether or not a particular item should confidently be categorized as metaphor (Cameron 1999a, 1999b), and the distinction between metaphor and metonymy has been found to be particularly blurred (Dirven 2003; Barnden 2010). However, our qualitative analysis of the students' responses at times allowed us to tell whether they were interpreting expressions as metaphor or metonymy. We also included between three and six distractors (e.g. technical terms, infrequent lexical items) for each lecture.

We had three students, whose first language was English, complete the metaphor interpretation task under similar conditions. This validated our own ideas of what would be reasonable to expect as valid responses under the time constraints of Activity 2 (see research question 3 below).

Examples of items selected for data collection Activity 2

Metaphorical items chosen for Activity 2 included ones that contributed to content teaching (e.g. *it started life* as a sign) and ones that offered content evaluation (e.g. climate change *demands* a global response). We also included some that were used for agenda management, such as signals on the direction or interrelationship of topics. These included Lecturer B's *bridge into*, signaling a gradual topic shift and Lecturer D's *that's what we're going back to,* signaling a return to a previous theme.

Our examples of conventional metaphor included phrasal verbs (e.g. *taking over; wrap up*) and highly commonplace metaphorical expressions (e.g. *give rise to; major shakeup; hot issue*). Some of the latter could count as metaphorical idioms. Further, some conventional metaphors that we included were multifaceted—for example, *explosion of interest*: this phrase connotes not just an increase, but also that it is both a *large* one and a relatively *quick* one. This afforded an opportunity to learn how closely the students could capture such multiple facets in their explanations. We also chose certain metaphorical uses of spatial prepositions that we thought might cause difficulty, such as in

inside a language. Towards the other end of the spectrum, we included relatively novel metaphors. One of the most novel examples was *the terrain of the verbal*.

We also included examples of personification, in keeping with Low *et al*'.s (2008) observation that it accounts for much of the recurrent metaphor in university lectures. In the Lecture B transcript, for example, personification was heavily used to give human agency to economic forces (e.g. *let the market do it*). A variation of personification—oracy—was also seen in the attribution of speech to theories (e.g. *The basic theory says that*. . .).

As indicated, these items were predominantly metaphorical, although some metonymy was also present, as in Lecturer B's *if you sat down a group of people (to try and plan what society needed),* used in relation to central economic planning. While the item could be taken literally to refer just to getting people to take a seat, it could also be taken metonymically to evoke many other aspects of a meeting scenario. References to metaphorical items in this article should be understood to include these occasional metonymies. More examples of the metaphorical items selected along with the metaphoric densities of each of the lectures can also be found on the companion website.

RESULTS[2]

Research question 1: of those language items that students found difficult to understand, what proportion were metaphorical?

For each lecture, our first step in investigating research question 1 was to list all of the items (words or phrases) identified by some students as problematic (i.e. difficult to understand) during Activity 1 for that lecture. Secondly, we scrutinized each of these items for metaphoricity, using the MIP for non-clear-cut cases and treating multiword items as metaphorical if they contained any metaphorical words. Thirdly, we calculated the proportion *M/P* for each student, where *P* is the number of problematic items identified by the student and *M* the number of metaphorical items among those problematic items. Fourthly, we averaged the individual students' *M/P* values to calculate an overall *M/P* value for the lecture.

After these calculations were performed for each lecture, our final step was to calculate the average of all the individual students' *M/P* values across all the lectures. The results of these calculations are shown in Table 2.

As indicated in Table 2 metaphor was indeed a problem for the students. Over all lectures, it accounted for ∼42 per cent of students' difficulties. Although this article is focused on the aggregate results across all lectures, we can make some preliminary conjectures about the differences between lectures, pending statistical analyses in future studies. The wide range of average *M/P* values for the individual lectures (21–60 per cent) appears unrelated to differences in metaphor densities, as such differences were minimal (ranging from 3.6 to 5.2). Rather, we conjecture that the variation across lectures

Table 2: Proportions of metaphorical items among students' problematic items in Activity 1

Lecturer	No. of students	Proportion of metaphor among problematic items[a] (%)
A	14[b]	20.8
B	4	57.5
C	12	59.5
D	12	45.4
Overall average[c]		42.4

[a]For each lecture: average of all individual-student *M/P* values (see text).
[b]Only 14 of the 18 students completed Activity 1. The others arrived late.
[c]Weighted average of the four per-lecture values, the weights being the corresponding numbers of students. Equivalent to the average of all individual-student *M/P* values irrespective of lecture.

reflected differences in the ways in which each lecturer chose to teach and comment on the subject content, as well as to scaffold the talk through agenda-management markers. Some of these differences are as follows.

In Lecture A on semiotics, where metaphor accounted on average for only 20.8 per cent of students' difficulties, it appears that students' attention was concentrated on subject-specific vocabulary and general lexical items that they did not understand. Non-metaphorical linguistic terms such as *semiotics, phonology, structuralism,* etc. were central to the lecture topic and used repeatedly and these were the types of words that often caused difficulty. Many of the other items reported as problematic were higher order words such as *finiteness, critique,* and *deterministic.* The metaphorical items that students reported as difficult to understand (e.g. *fused, array, stand out*) mostly occurred in the context of examples.

The average *M/P* proportion (57.5 per cent) for Lecture B must be treated cautiously because only four students attended the session. Nevertheless, we found the data, qualitatively, to be instructive. Much of the metaphor in this lecture, including the items reported as problematic, consisted of technical or quasi-technical economics terms (e.g. *the invisible hand*) that could be regarded as conventional metaphors (though they might be approached as novel metaphors by students who are relatively new to the subject, and/or who lack English as their first language).

In contrast, in Lecture C on the history of English dialects, although the average *M/P* proportion (59.5 per cent) was the highest of the four lectures, few of the problematic items were technical terms. Most were general lexical items (e.g. *fuelled by, lay down, melting pot*) used metaphorically to explain and exemplify the subject content. We noticed a small set of metaphorical words (e.g. *far-flung, foolproof*) that were cited by numerous students as problematic.

In Lecture D, the moderately high average *M*/*P* proportion (45.4 per cent) may be due partly to his greater use of novel metaphors relative to the other lecturers. In particular, Lecturer D often used figurative language to signpost an aside or even as part of the aside itself. While asides increase the organizational complexity of a lecture, their intent, as Strodt-Lopez (1991) states, is to 'increase the understandability of lectures'. However, the data collected from Lecture D suggest that they may sometimes have had the opposite effect.

In sum, across the lectures overall, an average of ~42 per cent of the words or phrases that a student found difficult to understand were, in fact, metaphorically used items. The reason for such metaphor use was to enhance comprehension of the lecture topic, through explanation, exemplification, evaluation, and so forth. In the three lectures with the highest proportions of such problematic metaphorical items (Lectures B–D), Lecturer B used more quasi-technical whereas the other two lecturers used more everyday, colloquial expressions. The fact that such a large proportion of the items presenting difficulties to the students across all four lectures were associated with these uses of metaphor suggests that they were not able fully to benefit from their attendance at these lectures.

Research question 2: of those language items that students found difficult to understand *despite being composed of familiar words*, what proportion were metaphorical?

In investigating research question 1, we discovered that students had difficulty in understanding certain metaphorical items in the lectures. However, we did not initially determine whether the students were simply unfamiliar with some of the words in these items. Research question 2 was aimed at making this distinction, which is important as ~85 per cent of the problematic items contained one or more unfamiliar words. We deemed that students were 'unfamiliar' with a particular word if they indicated never having encountered it, in any context.

As indicated earlier, in Activity 1 the students marked up the lecture transcript twice, in the second round, highlighting any words with which they were unfamiliar (as defined above). For research question 2, we excluded all problematic items containing such marked-up words; this modified an individual student's *P*-value (see discussion of research question 1) to count only those items whose usage in the lecture the student had found problematic despite its being lexically familiar (i.e. the student had encountered all of the words before—note that lexically familiar does *not* mean that the item as a whole is familiar to the student). Using our metaphoricity judgements from the analysis for research question 1, we then recalculated the *M*-value for each student, that is, the number of remaining *P* items that were metaphorical. These were the items that we suppose to have been problematic to a student specifically because of their metaphoricity.

Table 3: Proportions of metaphorical items among students' lexically familiar problematic items in Activity 1. Also shows how many of the problematic items were lexically familiar

Lecturer	No. of students	Proportion of problematic items that were lexically familiar[a] (%)	Proportion of metaphor among lexically familiar problematic items[b] (%)
A	14[c]	18.1	29.9
B	4	23.1	73.2
C	12	15.6	37.8
D	12	8.7	46.5
Overall average[d]		15.2	41.0

[a]For each lecture: the average of all individual-student proportions of lexically familiar with respect to problematic items. Lexically familiar = the student is familiar with all the words in the item *individually*.
[b]For each lecture: the average of all individual-student *M/P* values after all problematic items that were not lexically familiar to the student in question were excluded.
[c]Only 14 of the 18 students completed Activity 1.
[d]Weighted average of the four per-lecture values, the weights being the corresponding numbers of students. This is equivalent to the average of all individual-student values irrespective of lecture.

Using the adjusted per-student *M/P* values, we then calculated the four lecture averages and the overall average in the same manner as for research question 1. The results are shown in Table 3.

The proportion of metaphorical items on these shortened lists of problematic items jumped considerably for Lectures A and B relative to Table 2, presumably because many of the lexically unfamiliar items were not metaphorical. (Again, numerical results for Lecture B should be viewed with caution, because of the small number of respondents.) In contrast, in Lecture C, the average *M/P* value dropped markedly. Most probably, this was due to the exclusion of certain metaphorical words that had earlier boosted the research question 1 results in Lecturer C's case because they had been cited repeatedly as problematic by numerous students. Now it is clear that the comprehension problem with these words was not their metaphoricity but simply their unfamiliarity to the students. However, the overall average *M/P* across the lectures, 41 per cent, was almost the same as that in Table 2.

The fact that such a large proportion of the problematic items that were made up only of familiar words were metaphorical suggests that metaphor is likely to have made a major impact on the students' global understanding of the lecture as a whole. With hindsight, it would have been good to measure their global understanding more explicitly via a series of comprehension questions.

Research question 3: what kinds of errors did the students make, and how often did they make them, when asked to explain the contextual meaning of metaphorical items in the lectures?

Our investigation of research question 3 was based on the Activity 2 data, namely, the students' written explanations of various metaphors as used in the lecture contexts. We wanted to understand what kinds of errors students made in interpreting various metaphors or metaphor types. Much of our work for research question 3 focused on developing criteria for classifying student metaphor interpretations either as valid or as invalid in various different ways.

Beginning with a sampling of the Activity 2 student responses, each researcher independently studied each response to determine if it would be valid from an expert English-speaker point of view and, if not, why not. Later, through meetings and emails, we reached consensus on our results, and developed a taxonomy of eight types of metaphor misinterpretation error. We added two further types of response, namely: (i) when the response was too poorly expressed to be classifiable under the eight misinterpretation types and (ii) when the student explained something other than the intended (i.e. targeted) metaphorical item. We then worked in pairs to code the remaining Activity-2 student responses according to these 10 error types, while also noting which responses were *valid* metaphor interpretations in our judgement, and which were *null*. A *null* response was a blank one, an 'X' (which students were asked to use to indicate non-understanding), or occasional other responses that we could not construe as an attempt at interpretation.

We allowed as *valid* those answers that we felt correctly explained a metaphorical item's meaning by using either what we judged to be literal terminology or sufficiently different metaphorical terminology, in recognition of the fact that it is often difficult to explain a metaphor without recourse to another metaphor. By 'sufficiently different' we mean that the student used a metaphor with a qualitatively different source domain from that used in the actual metaphor. For example, one metaphorical item in Activity 2 for the Lecturer C lecture was 'you've got it'. We accepted literal responses such as 'you understand it' as being valid. We would also have accepted an example like 'you've seen the point' as a valid response using sufficiently different metaphorical terminology because it uses the source domain of vision. However, we judged the response 'you received it,' which one student gave, as being too close to the original source domain. But in cases where no alternative source matter would have been readily available to an expert English speaker in our judgement, we were prepared to allow explanations using the original source subject matter.

With regard to responses that were neither *null* nor *valid*, the 10 inadequate-response categories that we developed are identified below. These categories are each accompanied by one or more examples of the form I → R, where I is one of the items used in Activity 2, and R is an

actual response by a student. Within each I, the intended (i.e. targeted) metaphorical part is underlined (both in the text below and in the material presented to the students in Activity 2). Additional comments are included in brackets.

(Unclassified)
Unclassifiable because of poor expression

- Visual messaging *can get us beyond the barriers* of ordinary languages → *take some off* beyond the problems of ordinary languages (We were not confident we knew what the student meant by 'take some off'.)

(Neglected)
Not explaining the intended metaphorical part

Something else (which may or may not be metaphorical) is interpreted (perhaps correctly) instead of the intended metaphorical part, which appears essentially unchanged in the response or is apparently neglected.

- *basic gene pool → the fundamental or crucial gene pool* (The intended part is preserved without explanation in the answer.)

(Wrong grammar)
Apparently wrong lexico-grammatical analysis of the intended metaphorical part

The *Neglected* problem is avoided, but the intended metaphorical part's lexico-grammatical make-up is apparently analysed in a way that is clearly inappropriate. We include here cases of the participant not recognizing a fixed phrase or compound word.

- *foolproof → stupid evidence* (The fixed meaning of a compound word is not recognized.)

(Mistargeted)
Interpreting the intended metaphorical part as being about the wrong aspect of the described situation

- *the community is kept at a particular uniform level → common way of speaking* ('Uniform' is misattributed to a way of speaking instead of the social environment.)

(Stay in source)
Staying (too much) within the source subject matter

The interpretation stays within the source subject matter of the original metaphor, even though this could readily have been avoided (by using target-domain terminology or in the source terminology of a distinctly different metaphor). This response category applies irrespective of whether the re-expression is reasonable or not in its own terms or whether or not it is correct in context.

- *I should* confess *→ I want to say something that may be a secret*

(Unmotivated)

Apparent lack of metaphorical motivation based on the intended metaphorical part

An interpretation is given of the intended metaphorical part, but it is wrong in context and it is not readily apparent that it is metaphorically derived from any meaning of the intended metaphorical part.

- *have worked on you since... → interrupted your life*
- *strictly → confidently*

(Wrong/commonplace)

Inappropriate metaphorical interpretation: commonplace case

An interpretation is given of the intended metaphorical part which is close to a commonplace one in English. However, the interpretation is still inappropriate in context.

- *some point over the next week → some interesting subject over next week* ['Point' can mean '(main) topic' and therefore (roughly at least) 'subject', but this meaning is inappropriate in the context because a temporal sense of 'point' was intended.]

(Wrong/non-commonplace)

Inappropriate metaphorical interpretation: non-commonplace case

An interpretation is given of the intended metaphorical part. While it is not a commonplace metaphorical interpretation, the interpretation is nevertheless apparently motivated metaphorically. However, the interpretation is still inappropriate in context.

- *stem from X → seem clearly different from X* (The metaphor of a stem coming out of a branch is being non-standardly interpreted as conveying a different direction of development.)

In this response category, the apparent metaphorical motivation may not even be a workable one in English in any context, so transfers of metaphorical motivation from other languages are included here.

(Overspecified)

Over-specification

The interpretation is valid except for being too narrow.

- *social network → social conversations* (These are just part of being in a network.)
- *acquire their own government → set up their own government* ('Set up' implies 'create' or 'establish', which is more specific than 'acquire'.)

(Underspecified)

Under-specification

The interpretation is valid except for being too broad.

- *confess → say* ('Confess' involves more than 'saying'.)

- *explosion → its usage has increased a lot* (This does not capture sense of suddenness.)

We regard all these response types except for *unclassified*, *neglected*, and *null* as clear attempts to interpret the intended metaphorical parts. We therefore refer to these responses, together with *valid* ones, as *applicable*. In contrast, in *unclassified, neglected,* and *null* responses, it seems that the students have not addressed the intended metaphorical parts (whether through lack of time, misunderstanding the instructions, or for some other reason) or it is unclear whether they have or not. We therefore refer to these responses as *not applicable (N/A).*

In most cases, only one category was assignable to each response. The exception was that *wrong grammar* could be assigned in concert with a type lower in the list. In such cases we assigned only one response type to each item, reflecting what we considered to be the primary one. We then totalled all the response types per lecture and calculated the percentage share of each type per lecture. We also calculated an overall figure across all lectures. The results are shown in Table 4.

As can be seen, the most common of the 10 types in the list above, based on the overall figure across the lectures, were, in decreasing order, *underspecified*, *neglected*, *unmotivated, stay in source,* and *overspecified*. Taken together, these error types were the most prevalent ones across the lectures, accounting for ∼31 per cent of the total. Originally we had included a category called 'cultural and linguistic transfer' as we had expected this to be an important factor influencing the interpretations offered by our students. However, when we analysed their interpretations we found no clear evidence of any L1 influence and discussions with the students themselves yielded no examples of L1 cultural influence. We therefore decided to eliminate this category. The metaphors investigated in this study were simply those that arose spontaneously in the lectures attended by the students and it may simply have been the case that none of these metaphors had false friends in the students' L1. Our inability to identify L1 influence may constitute a shortcoming of our study as it has been convincingly argued by a number of researchers (e.g. Lantolf and Thorne 2006) that the conceptual metaphors in a student's first language will affect their ability to grasp and to appropriate different conceptual metaphors underlying the target language. We can only assume that such L1 influence operates at a deeper, more conceptual level that was not accessible through our chosen methodology. Cultural/linguistic transfer is a phenomenon that could usefully be explored in a more controlled, in-depth qualitative study focusing specifically on metaphors whose wording corresponds in a misleading way to metaphorical wording in an L1.

Extent to which responses were (broadly) valid or not

After our detailed analysis of metaphor interpretation errors, we turned our attention to the proportions of *valid* versus non-valid (*invalid*) responses.

Table 4: Prevalence of the 10 response types (excluding null and valid) in Activity 2. For each lecture and code, the number shown is the percentage of responses (over all students in Activity 2 for that lecture) that were given that code, out of the responses given one of the 10 codes

Lecturer	*N*[a]	*U*	Neg	Wrong gramm	Mis-targ	Stay in source	Un-motiv	Wrong/ common	Wrong/ non-common	Over-spec	Under-spec
A	18	5.1	20.8	3.2	5.1	6.0	18.1	7.9	2.3	17.6	13.9
B	4	9.5	2.4	4.8	0.0	26.2	4.8	0.0	2.4	4.8	45.2
C	12	6.8	24.0	3.4	4.8	11.6	14.4	15.1	4.1	11.0	4.8
D	12	3.1	9.0	9.0	3.6	13.5	7.2	4.9	13.0	5.8	30.9
Overall[b]		5.1	16.1	5.4	4.1	11.3	12.4	8.0	6.5	11.0	19.9

[a]*N* = number of students.
[b]For each code: the percentage of responses (over all students in Activity 2 for any lecture) that were given that code, out of the responses that were given one of the 10 codes shown.
Code abbreviations: U = unclassified, Neg = neglected, gramm = grammar, targ = targeted, motiv = motivated, common = commonplace, spec = specified.

Table 5: Valid and problematic metaphor interpretations in Activity 2, under the generous scoring criterion

Lecturer	Null (%)	Broadly valid (%)	Not broadly valid (though applicable) (%)	Not broadly valid (though applicable)/all applicable (%)	Not applicable (%)	Problematic (%)
A	9.5	64.8	16.0	19.8	19.3	35.2
B	16.0	63.0	16.0	20.3	21.0	37.0
C	32.7	41.0	16.7	28.9	42.3	59.0
D	42.8	34.6	18.3	34.5	47.1	65.4
Overall[a]	24.8	50.6	16.8	26.1	32.7	49.4

[a]Average of the per-lecture values, weighted by the number of students at corresponding sessions.
Problematic = not broadly valid or not applicable.
All percentages shown are proportions with respect to all responses except where shown as being with respect only to all applicable responses.

We were interested in assessing the extent to which the participants were unable to explain the metaphors selected. As it is sometimes very difficult to explain the meanings of metaphors, particularly in a foreign language (which most of the students chose to do) we adopted a generous scoring procedure which involved calculating for each lecture the proportion of responses that could be classified as *broadly valid*. This included all responses coded as *valid*, *overspecified* or *underspecified*. Table 5 shows that the average proportion of problematic items is 49.4 per cent, that the proportion (out of *all* responses) of *not broadly valid though applicable* items is 16.8 per cent, and, more tellingly, that the proportion of *not broadly valid* items out of all *applicable* responses is 26.1 per cent. We also adopted a stricter scoring criterion in which over- and under-specified items were marked as incorrect. This revealed an even higher error rate of 42.8 per cent (*invalid* out of all *applicable)*. Details are provided on the companion website. In contrast, the three students whose native language was English gave on average ~94 per cent of the metaphorical items a *broadly valid* interpretation.

Research question 4: to what extent were students aware of their errors in attempting to understand metaphorical items in the lectures?

Research question 4 addresses the 'silent' aspect of metaphor misunderstanding—the extent to which students are unaware of a comprehension difficulty, and therefore leave the lecture hall with uncorrected notions regarding the lecture content. Because our participants could only report difficulties of which they were aware, we needed to devise an indirect way of measuring their

Table 6: Extent to which a metaphor found to be interpreted in a not broadly valid way in Activity 2 was also identified as being problematic by the same student in Activity 1

Lecturer	No. of students	*Not broadly valid though applicable* responses in Activity 2 that were also identified in Activity 1[a] (%)
A	14[b]	5.9
B	4	0.0
C	12	5.8
D	12	2.1
Overall[c]		4.2

[a]For each lecture: average over the students in Activity 2 for that lecture of: the proportion of the metaphorical items given an *Not broadly valid though applicable* interpretation in Activity 2 by a student that that student also identified as problematic in Activity 1 for that lecture.
[b]Only 14 of the 18 students completed Activity 1.
[c]Average of per-lecture values weighted by the numbers of students. Equivalent to average over all students and lectures of the per-student proportions.

unawareness of problems in understanding. This we did by bringing together the data from the earlier parts of our study, as described below.

We began by listing, for each student and lecture, all of the metaphorical items in Activity 2 for that lecture where the student's response was *not broadly valid though applicable*. We then identified which of these invalid Activity 2 items had been among the problematic items the student had identified earlier in Activity 1. In such a case, we reasoned that the student was aware of a comprehension problem, even if he/she did not fully recognize the metaphorical aspect involved. In contrast, if an invalid Activity 2 item had not come up during Activity 1, we reasoned that this represented a metaphor comprehension problem of which the student was unaware, that is a misinterpretation in the sense introduced in the Introduction. The proportion of the Activity 2 items receiving an *not broadly valid though applicable* response by a student that were not reported by that student as problematic in Activity 1 became our main quantitative measure of the extent to which the student was unaware of their metaphor misunderstanding in the lecture.

On this basis, Table 6 shows the resulting proportions per lecture, and an overall proportion. The results show that, in all four lectures, the students appear to have been largely unaware of the problems with their metaphor interpretations. Of the metaphorical items for which the students gave *applicable* but not broadly valid interpretations, only 4.2 per cent had been flagged in Activity 1 by them as posing difficulties.

These findings are particularly interesting against the findings to research questions 1 and 2. There, we found that many of the items that the students

had identified as problematic even though composed of familiar words were in fact metaphorical. As can now be seen, even the large number of recognized problems did not come close to revealing the full extent of the metaphor misunderstanding actually occurring at these lectures.

DISCUSSION

This study has investigated some of the ways in which figurative language can cause comprehension difficulties for university students whose first language is not English. A study was conducted of four lectures attended by 20 International Foundation Programme students whose first language was not English, at a British university. Through four post-lecture sessions with the students, a wealth of data were collected on: (i) the extent to which items that were problematic for students despite being lexically familiar were metaphorical, and (ii) the extent to which, and the ways in which, the students had difficulty with the figurative language used. One main finding on (i) was that overall ~41–42 per cent of the lexically familiar but problematic items were metaphorical. A main finding on (ii) was that, for the lectures overall, even when generous scoring criteria were used, the students failed to give a *valid* interpretation of about 26 per cent of the intended metaphorical items that they attempted to interpret. However, they were aware of having a problem with only ~4 per cent of their poorly interpreted items.

These findings are worrying because, ordinarily, the use of metaphor is a valuable teaching tool. Lecturers use it to explain, clarify, summarize, evaluate; to remind or challenge; and above all, to make their lectures easier to understand. It does, however, appear to present problems to international students whose first language is different from that of the lecturer.

It is important to consider at this point the extent to which our method genuinely gets at the difficulties that metaphor presented to these students. First, there was a 2-week delay between hearing the whole lecture and the students being presented with parts of it, which may have made the task more difficult for the students and certainly reduced the authenticity of the study. Secondly, as we have already pointed out, the fact that the students were asked to explain the metaphors may have raised the problem that they lacked the vocabulary in English to explain their ideas properly (although it must be remembered that we did give them the opportunity to answer in their own language). These constitute weaknesses of the study, as they may have rendered the task artificially difficult.

On the other hand, there are other features of our research design that actually rendered the task easier for these students in some ways. First, our use of short extracts minimized the interaction of figurative difficulty and exhaustion and meant that the students no longer had to cope with the complexity of an hour's lecture content. And secondly, the use of transcripts minimized difficulties due to speed of delivery, acoustic problems, and elision.

The fact that our students still experienced difficulties with metaphor even in this situation strengthens our argument.

More research is needed to fully investigate the problems explored in this article, and to help guide practitioners who are working with English for Academic Purposes (EAP) students on their metaphor interpretation strategies. We believe that our methodology, particularly our newly developed scheme of metaphor misinterpretation errors, may be useful for such future teaching and research efforts. However, given the variability of the data collected in the current study, due in part to the small sample size, larger scale studies are needed so that statistically reliable results can be obtained. Our error scheme also needs further testing and refinement. Further research could also be usefully conducted into the kinds of problems caused by different broad types of metaphors (e.g. conventional metaphor, novel metaphor, personification).

Another future direction would be to continue with ongoing research to determine what variations of figurative language use in lectures might be attributable to such parameters as subject area and/or lecture style. Developing ranges of variability for lecture metaphor densities would also help establish a background against which the representativeness of the results of particular studies can be better assessed.

To summarize, while we believe that both the methods and results of this study are illuminating, more research needs to be done to understand how best to help international students whose first language is not English to develop their metaphor comprehension skills, particularly in the context of academic lectures. It would also be useful for lecturers to be aware of the range of potential difficulties that metaphor presents to such international students, and to take measures to ensure that key metaphors used in their lectures have in fact been understood by all. If this can be done, metaphor can become, for them, the same effective tool of intellectual communication that it is for others.

SUPPLEMENTARY DATA

Supplementary material is available at *Applied Linguistics* online.

NOTES

1 In Lecture A, one student gave six responses in both English and Spanish, which he then translated himself into English. The translations were checked by Spanish-speaking members of the research group (Littlemore and Koester). In Lecture B all students answered in English. In Lecture C, three students gave a total of 10 responses in their native language. One provided an English translation as well; the other two wrote the native language words only, which were subsequently translated by the Chinese-speaking member of the research team (Chen). In Lecture D, three students gave a total of seven responses in their own language. One, whose first language was

Russian, provided an English translation as well and the Russian words were the exact equivalent of what was in the text. The other two only wrote in their native language, and their answers were subsequently translated from Chinese or Spanish by the researchers.

2 Some preliminary and partial findings from this study were published in Littlemore *et al*. (2010).

REFERENCES

Barnden, J. A. 2010. 'Metaphor and metonymy: Making their connections more slippery,' *Cognitive Linguistics* 21/1: 1–34.

Cameron, L. 1999a. 'Operationalising 'metaphor' for applied linguistic research' in L. Cameron and G. Low (eds): *Research and Applying Metaphor*. Cambridge University Press, pp. 1–29.

Cameron, L. 1999b. 'Identifying and describing metaphor in spoken discourse data' in L. Cameron and G. Low (eds): *Research and Applying Metaphor*. Cambridge University Press, pp. 105–34.

Cameron, L. 2003. *Metaphor in Educational Discourse*. Continuum.

Cameron, L., R. Maslen, Z. Todd, J. Maule, P. Stratton, and **N. Stanley.** 2008. 'The discourse dynamics approach to metaphor and metaphor-led discourse analysis,' *Metaphor and Symbol* 24/2: 63–89.

Corts, D. and **H. Pollio.** 1999. 'Spontaneous production of figurative language and gesture in college lectures,' *Metaphor and Symbol* 14/2: 81–100.

Dirven, R. 2003. 'Introduction' in R. Dirven and R. Pörings (eds): *Metaphor and Metonymy in Comparison and Contrast*. Mouton de Gruyter, pp. 1–38.

Drew, P. and **E. Holt.** 1998. 'Figures of speech: figurative expressions and the management of topic transition in conversation,' *Language and Society* 27: 495–522.

Lantolf, J. P. and **S. L. Thorne.** 2006. *Sociocultural Theory and the Genesis of Second Language Development*. Oxford University Press.

Littlemore, J. 2001. 'Metaphor as a source of misunderstanding for overseas students in academic lectures,' *Teaching in Higher Education* 6/3: 333–51.

Littlemore, J., P. Tang, P. Chen, J. Barnden, and **A. Koester.** 2010. 'The role of figurative thinking in gaining access to discourse communities: A report on two case studies' in S. De Knop, F. Boers, and T. De Rycker (eds): *Exploring the Lexis-Grammar Continuum in Second Language Pedagogy*. Mouton de Gruyter, pp. 189–211.

Low, G., J. Littlemore, and **A. Koester.** 2008. 'The use of metaphor in three university academic lectures,' *Applied Linguistics* 29/3: 428–55.

McCarthy, M. 1998. *Spoken Language and Applied Linguistics*. Cambridge University Press.

Pragglejaz Group. 2007. 'A practical and flexible method for identifying metaphorically-used words in discourse.' *Metaphor and Symbol* 22/1: 1–40.

Steen, G. J. 2007. *Finding Metaphor in Grammar and Usage: A Methodological Analysis*. John Benjamins.

Strodt-Lopez, B. 1991. 'Tying in all in: asides in university lectures,' *Applied Linguistics* 12/2: 115–40.

Thompson, S. 1994. 'Frameworks and contexts: a genre-based approach to analyzing lecture introductions,' *English for Specific Purposes* 13/2: 171–86.

NOTES ON CONTRIBUTORS

Jeannette Littlemore is a Senior Lecturer in the Centre for English Language Studies at the University of Birmingham. Her current research interests lie in the acquisition and use of figurative language by second language learners. Her publications include: Littlemore, J. and G. Low. 2006. *Figurative Thinking and Foreign Language Learning*. Palgrave Macmillan. Littlemore, J. 2009. *Applying Cognitive Linguistics to Second Language Learning and Teaching*. Palgrave Macmillan. *Address for correspondence*: The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. <*j.m.littlemore@bham.ac.uk*>

Phyllis Trautman Chen has an MA in Applied Linguistics from the University of Birmingham and an MA in East Asian Studies from Yale University. She is currently teaching Mandarin Chinese at the Bronx High School of Science, New York. *Address for correspondence*: The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. <*phyllistchen@yahoo.com; chen1@bxscience.edu*>

Almut Koester is a Senior Lecturer in the Department of English at the University of Birmingham. Her research is in the area of spoken workplace discourse, and she is particularly interested in the function of interpersonal features of language, such as idioms, metaphor, and hedges in workplace interaction. Relevant publications include: Koester, A. 2004. *The Language of Work.* Routledge. Koester, A. 2006. *Investigating Workplace Discourse*. Routledge. *Address for correspondence*: The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. <*a.j.koester@bham.ac.uk*>

John Barden is a Professor of Artificial Intelligence in the School of Computer Science at the University of Birmingham. His research is within Artificial Intelligence and focuses on the understanding of metaphorical utterances and the representing and reasoning about mental states of agents. Relevant publications include: Barnden, J. A. 2007. 'Metaphor, semantic preferences and context-sensitivity' in K. Ahmad, C. Brewster, and M. Stevenson (eds): *Words and Intelligence II: Essays in Honor of Yorick Wilks*. Springer, pp. 39–62. Barnden, J. A. 2010. 'Metaphor and metonymy: Making their connections more slippery.' *Cognitive Linguistics* 21/1: 1–34. *Address for correspondence*: The University of Birmingham, Edgbaston, Birmingham B15 2TT, UK. <*j.a.barnden@cs.bham.ac.uk*>