

'...post-colonialism, multi-culturalism, structuralism, feminism, post-modernism and so on and so forth': Vague category markers in academic discourse, a comparative analysis

Steve Walsh, Anne O'Keeffe & Michael McCarthy

ABSTRACT

The use of vague language is one of the most common features of everyday spoken English. Speakers regularly use vague expressions to project shared knowledge (e.g., *pens, books, and that sort of thing*) as well as to make approximations (e.g. around sevenish, he's sort of tall). Research shows that many of the most common single word items in a core vocabulary form part of vague language fixed expressions (e.g. *thing in that kind of thing*). This paper will address the use of vague language in a new corpus of academic English, the Limerick-Belfast Corpus of Spoken Academic English (LIBEL case). The LIBEL corpus consists of one million words of spoken data collected in two universities on the island of Ireland, one in the Republic of Ireland and one in Northern Ireland. Analysis of the LIBEL corpus will identify forms and functions of vague language in an academic context and these findings will then be compared with two corpora of everyday spoken language from the Republic of Ireland and the United Kingdom, namely the Limerick Corpus of Irish English (LCIE) and the Cambridge and Nottingham Corpus of Discourse in English (CANCODE). Cross-corpora comparison allow us to look at how forms and frequencies of certain vague language expressions vary across casual and formal/institutional contexts. Within the academic data we build on Walsh's work (see for example Walsh 2002, 2006), we will also show how vague language use is relative to mode of discourse at any given stage of classroom interaction. We suggest that these qualitative differences are a valuable means of understanding the complex relationship between language and learning.

Suggested short title for running head: Vague category markers in academic discourse

'...post-colonialism, multi-culturalism, structuralism, feminism, post-modernism and so on and so forth': Vague category markers in academic discourse, a comparative analysis

1. Introduction: vague categories

Degrees of variation exist in how vague language is defined. Channell (1994) restricts it 'purposefully and unabashedly vague' uses of languages while Franken (1997) distinguishes between 'vagueness' and 'approximation'. Zhang (1998) makes a case for four separate categories: 'fuzziness', 'generality'; 'vagueness' and 'ambiguity'. Chafe (1982) puts vagueness and hedging in the same category of 'fuzziness' all of which are seen as 'involvement devices' more prevalent in spoken rather than written language. The notion of vagueness as an involvement device is consistent with the view that vague language as a core feature of the grammar of spoken language (Carter and McCarthy 1995, 2006; McCarthy and Carter 1995). As Carter and McCarthy (2006) note, vague language is a strong indicator of assumed shared knowledge which marks in-group membership insofar as the referents of vague expressions can be assumed to be known by the listener. This is consistent with Cutting (2000), who illustrates how discourse communities use vague language as a marker of in-group membership. The interactive aspect of vague language is important to our focus in this chapter where we examine the use of vague language in the learning context of university discourse. In this domain, the use of vague language is part of meaning making within specific learning contexts or *modes* (see Walsh, 2006: 111).

We will focus on one type of vague language, namely vague category markers (hereafter VCMs). These non-lexicalised categories are created within interactions, at the moment of speaking. The categories contain exemplars followed by a vagueness tag (*and so on, and that kind of thing, et cetera, and things like that*) and the listener(s) are expected and assumed to fill in, or implicitly understand the reference. Here is an example taken from a drama lecture in the Limerick Belfast Corpus of Academic Spoken English (LIBEL CASE¹) (see details below section 3):

1)

... And I suppose my understanding of critical theory and critical aah critical studies I suppose as such emanate from or are the key social critiques of our time which have emanated from the work of the Frankfurt School. So at the moment its you know ahh critical theory is dominated by ideas of **post-colonialism multi-culturalism structuralism feminism post-modernism and so on so forth.**

Here the exemplars are: *post-colonialism multi-culturalism structuralism feminism post-modernism*, and the tag which creates the VCM is: *and so on so forth.*

Here is an example from a corpus of casual conversation (the Limerick Corpus of Irish English, LCIE, see below). Here friends are chatting. Speaker (1) creates a VCM but the listener does not understand the exemplar. Hence the category is not created and needs further explanation. In the process of explanation, another VCM is created:

2)

<\$6>: He just made up words like he just made up I don't know what.

<\$1>: Is that not artistic license like? amm **coinage and stuff like that?**

<\$6>: What?

<\$1>: Coinage.

<\$6>: What's coinage?

<\$1>: **When you are writing poetry and stuff** you can make up your own words.

<\$4>: Yeah I mean yeah.

<\$6>: Like say sarcasamistic like?

<\$1>: Yeah <laughs> you are a poet and you don't know it my friend?

<\$5>: Ah snozberry.

<\$1>: Yeah.

<\$4>: Fantastic.

This is a good example of how meaning is negotiated interactively within a conversation. The first VCM which speaker 1 uses over-extended the range of assumed shared

knowledge between the speakers by using the exemplar *coinage*. The second VCM which she creates uses a much more general exemplar, *poetry*, which is obviously within the range of shared knowledge of the group.

2. Previous research into vague categories

Vague categories can be divided into lexicalised and non-lexicalised types. Lexicalised categories are those which provide superordinates or prototypes encoded as a single, lexical item, for example *bird*, *furniture*, *machinery*. Until recently most research into the nature of categories has been concerned with these lexicalised categories within the field of semantics, see in particular the work of Rosch and her associates (Mervis and Rosch, 1981; Rosch, 1978; Rosch et al. 1976), who demonstrated that the categories they studied had a graded structure and that at the centre of each category was a prototype that exhibited the highest concentration of characteristic properties compared with members at the periphery which contained fewest characteristic properties.

Non-lexical categories are ad hoc rather than prototypical. The concept is attributed to the work of Barsalou (1983, 1987), though links may be seen in the work of Cruse (1986) on what he called *lax hyponymy* (the non-institutionalised arrangements of items into instantial categories at the time of speaking). The question as to whether categories are stable or subject to change is addressed in particular by Barsalou (1983 and 1987), who talks about the dynamic nature of ad hoc category formation, for example *places to look for antique desks*. In such examples, categorisation is non-lexicalised and without clear boundary, challenging the notion that categories are stable, easily recognisable and arrived at 'pre-textually' (after Overstreet and Yule 1997a). Overstreet and Yule (1997a) reflect that:

If only common (i.e. lexicalised) categories are studied then little insight will be gained into the discourse processes involved in categorisation when a single lexical item is not available to the discourse participants for the referential category.

Building on the ad hoc categories of Barsalou (1983), they stress the spontaneity of categorisation and the context-dependent nature of the categories themselves when one looks at examples from actual discourse as opposed to stylised examples. Overstreet and Yule (1997a: 87) suggest a continuum from lexicalised to non-lexicalised categories based on the degree to which categories are: a) conventionally and linguistically established and b) constrained by contextual factors.

In the literature, the tags which help create these ad hoc categories go by different terms such as: ‘general extenders’ (Overstreet and Yule 1997a, 1997b) ‘generalized list completers’ (Jefferson 1990); ‘tags’ (Ward and Birner 1992) ‘terminal tags’ (Dines 1980; Macaulay 1991); ‘extension particles’ (DuBois 1993), ‘vague category identifiers’ (Channell 1994, Jucker, Smith and Lüdge 2003) and vague category markers (O’Keeffe 2003, 2006; Evison, McCarthy & O’Keeffe In press). In this chapter we adhere to the term *vague category marker* (VCM).

The questions of interest for this paper are: do such phenomena manifest themselves in spoken academic discourse, and if so, to what ends, and do such phenomena differ from or resemble uses of vague language in everyday causal conversation? This last question is important, since special registers in spoken language are often best characterised by illuminating the degree to which they resemble or depart from the typical linguistic features of banal conversation. We enter this investigation via the notion of *classroom modes*, a set of ways of communicating between teachers and students which recur in the academic corpus, and which seem to have clear pedagogical foci in relation to overall goals in educational settings. The notion of modes is based on the work of Walsh (2006).

3. Classroom modes

In this section, a framework for analyzing third-level spoken academic discourse is presented and exemplified. The framework, SETT (Self-Evaluation of Teacher Talk, Walsh, 2006), emphasizes the fact that interaction and classroom activity are inextricably linked, and acknowledges that as the focus of a learning event (e.g. lesson/seminar/workshop) changes, so interaction patterns and pedagogic goals change. When language use and pedagogic purpose are considered *together*, different contexts emerge, making it possible to analyze the ensuing discourse more fairly and more objectively (see, for example, van Lier 1988; Seedhouse, 2004). Under this variable view of contexts (plural), learner and teacher patterns of verbal behaviour can be seen as more or less *appropriate*, depending on a particular pedagogic aim. Characterizing third-level teaching in this way is not intended to offer an all-encompassing description nor a means to 'code' interaction patterns. Rather, the intention is to offer a framework and a metalanguage which may be used to interpret interaction in the context of third-level classrooms.

Like other writers who adopt a variable view of classroom context, the SETT framework, presented below, also adopts a variable approach. Specifically, the design of the framework rests on four assumptions. Firstly, all classroom discourse is goal-oriented: the prime responsibility for establishing and shaping the interaction lies with the teacher; secondly, pedagogic purpose and language use are inextricably linked – it is impossible to consider one without taking account of the other; thirdly, any higher education classroom context is made up of a series of micro-contexts (termed *modes*) which are linked to the social, political, cultural and historical beliefs of the participants (c.f. Kumaravadivelu, 1999); fourthly, micro-contexts are co-constructed by teachers and students through their participation, through face-to-face meaning-making and through a process of 'language socialization' (Pavlenko and Lantolf, 2000).

A mode is defined as a 'classroom microcontext which has a clearly defined pedagogic goal and distinctive interactional features determined largely by a teacher's use of language'. (Walsh, 2006: 111). A modes analysis recognizes that understanding and

meaning are jointly constructed, but that the prime responsibility for their construction lies with the teacher.

The original SETT framework is based on a corpus of 14 English for Specific Purposes lessons, totalling approximately 12 hours or 100,000 words. The framework has since been applied to a much larger corpus of one million words of academic spoken English recorded in two universities in Ireland. This corpus, LIBEL CASE (Limerick and Belfast Corpus of Academic Spoken English), is composed of spoken academic data, collected at Queen's University Belfast, Northern Ireland, and University of Limerick, Republic of Ireland, from the following contexts: lectures, seminars, small group tutorials, oral presentations and conference papers. 50% of the corpus was collected in each site and its design matrix spans subject areas and colleges within the two institutions, so as to achieve internal comparability and overall representativeness (see www.mic.ul.ie/ivacs).

Based on the initial corpus findings, qualitative samples the data was analysed by working from concordance lines. In the qualitative stage a CA methodology was used, which centred on turn-taking mechanisms in relation to the perceived goal of the moment and the stated (written) lesson aims given by the teacher. Interaction patterns were found to vary according to instructional activity; for example, establishing procedures to complete an activity resulted in a very different pattern of interaction to that of open-class discussion. The different patterns manifested themselves in the turn-taking, sequence of turns and topic management. According to Heritage, interactants' talk is 'context-shaped' by a previous contribution, and 'context-renewing' by subsequent ones; understanding is indicated by the production of 'next' actions (1997: 162-3). In other words, participants both contribute to and demonstrate understanding of the interaction through the ways in which turns are managed. In this way, it is possible to characterize both the relationship between talk, and actions and assess the extent to which the 'talk-in-interaction' is appropriate to the shifting agenda and pedagogic goals of the moment.

Following this procedure, it was possible, by analyzing the corpus, to identify four patterns, four micro-contexts, called *modes*: managerial mode, classroom context mode,

skills and systems mode, materials mode. Each mode has distinctive interactional features and identifiable patterns of turn-taking related to instructional goals. While other modes could almost certainly be identified (depending on the specific context), these four are included as being representative of the interaction which takes place in the third level classroom, because they provide clear-cut examples of different types of interactional patterning and because they are intended to be used by teachers using samples of their own data as a means of raising awareness.

Heritage and Greatbach's (1991) notion of 'fingerprints' is helpful to the present discussion. In that study, the researchers identify a number of socially constructed contexts in different institutional settings which they term 'fingerprints' to differentiate interactional organisations from one work-place to another. Thus, the 'fingerprint' of a doctor's surgery will have a different exchange and participation structure to that of a solicitor's office. Here, we are proposing that each classroom mode has its own distinctive fingerprint, comprising pedagogic and linguistic features. Thus, the fingerprint of *classroom context* mode is markedly different to that of *managerial* mode; both are different again from *skills and systems* mode.

The four modes, together with their interactional features and typical pedagogic goals, are summarized in table 1 below:

Table 1: Classroom Modes (Walsh, 2006)

Mode	Pedagogic Goals	Interactional features
Managerial	<ul style="list-style-type: none"> • To transmit information • To organize the physical learning environment • To refer learners to materials • To introduce or conclude an activity • To change from one 	<ul style="list-style-type: none"> • A single, extended teacher turn which uses explanations and/or instructions • The use of transitional markers • The use of confirmation checks • An absence of learner

	mode of learning to another	contributions
Materials	<ul style="list-style-type: none"> • To provide input or practice around a piece of material • To elicit responses in relation to the material • To check and display answers • To clarify when necessary • To evaluate contributions 	<ul style="list-style-type: none"> • Predominance of IRF pattern • Extensive use of display questions • Content-focused feedback • Corrective repair • The use of scaffolding
Skills and systems	<ul style="list-style-type: none"> • To enable learners to produce correct answers • To enable learners to manipulate new concepts • To provide corrective feedback • To provide learners with practice in sub-skills • To display correct answers 	<ul style="list-style-type: none"> • The use of direct repair • The use of scaffolding • Extended teacher turns • Display questions • Teacher echo • Clarification requests • Form-focused feedback
Classroom context	<ul style="list-style-type: none"> • To enable learners to express themselves clearly • To establish a context • To promote dialogue and discussion 	<ul style="list-style-type: none"> • Extended learner turns • Short teacher turns • Minimal repair • Content feedback • Referential questions • Scaffolding • Clarification requests

Owing to the multi-layered, ‘Russian doll’ (Jarvis and Robinson, 1997: 225) quality of classroom discourse, any classification is not without its problems and the present one is no exception. Tensions between and within modes do exist: rapid movements from one mode to another, termed *mode switching*; brief departures from one mode to another and back again, henceforth *mode side sequences*; the fact that some sequences do not ‘fit’ into any of the four modes identified. These have all posed problems for description.

Moreover, the analysis is further complicated by the homogeneous and heterogeneous quality of classroom contexts (Seedhouse, 2004); within a mode, every interaction is both similar to other interactions (homogeneous) and yet a unique encounter (heterogeneous).

3. Data and methodology

We draw on three spoken language corpora, LIBEL, from an academic setting and two comparable corpora composed of casual conversation from Britain and Ireland. Table 2 summarizes these data.

Table 2 – Description of data used in the study

Corpus	no. of works	Description
Limerick-Belfast Corpus of Academic Spoken English (LIBEL)	500,000 words ²	<ul style="list-style-type: none"> • A corpus of lectures, small group tutorials, laboratories and presentations. • Collected in two universities on the island of Ireland: Limerick and Belfast³ • Data collected from common disciplinary sites: Arts and Humanities, Social Sciences, Science, Engineering and Informatics and Business.
Cambridge and Nottingham Corpus of Discourse in English (CANCODE)	5 million words in total, here a sub-corpus of 1 million words was selected	<ul style="list-style-type: none"> • Consisting of casual conversations between family and friends in Britain and Ireland. • Designed to reflect spoken genres, speaker relationships and context (see McCarthy 1998) • A one-million word sub-corpus was selected for use here. This comprises casual conversation between family and friends in Britain.
Limerick Corpus of Spoken English (LCIE)	1 million words	<ul style="list-style-type: none"> • Designed as a comparable corpus to CANCODE. • Consisting of casual conversations between family and friends in Southern Ireland.

In this paper we draw on two methodologies not always seen as complementary, corpus linguistics and conversation analysis. These have much to offer each other as they provide both quantitative and qualitative insights respectively (McCarthy and Carter 2002, O’Keeffe 2006, Walsh and O’Keeffe In press). Applied to these corpora, *Wordsmith Tools* software (Scott 1999) was used to produce word cluster (or chunks) frequency lists, that is to say, lists of recurrent strings of pre-selected extents (e.g. three-word clusters, four-word clusters). These quantitative data were sorted so as to identify VCMs in each dataset (LIBEL, CANCODE sub-corpus and LCIE). This process involved concordancing individual high-frequency chunks operating as VCMs, and extensive manual reading of sample files. When we look at the micro-contexts, or modes, we employ CA to help understand the ways in which vague language is manifested in each mode, and the contribution VCMs make to the enactment of the modes. A full account of the transcription conventions used appears in the appendix. Table 1 should be used as a reminder of the interactional features and pedagogic goals of each of the four modes.

4. Analysis

The quantitative findings based on the three corpora are illustrated below. These show the most common VCM forms and their frequencies in the three datasets. These forms came from cluster analyses using *Wordsmith Tools*.

Table 3 – VCM forms resulting for cluster analysis (normalised to occurrences per million words)

Form	CANCODE	LCIE	LIBEL
or something	717	590	0
and all	0	447	0
and stuff	326	0	0

and that	117	420	0
et cetera	0	0	229
and so forth	0	0	151
and all that	0	90	0
or something like that	88	49	71
and things like that	77	44	0
and so forth like that	0	0	59
and stuff like that	57	24	0
(and) that sort of thing	56	0	0
(and) that kind of thing	0	31	55
or anything like that	23	30	0
and so on and so forth	0	0	27
this that and the other	16	0	0
(and) all that sort of thing	9	0	0
all this kind of stuff	7	0	0
and all the rest (of it)	12	40	0
and all that sort of thing	8	0	0
and all this/that kind of stuff	12	0	0
and this that and the other	5	0	0
and so on	0	5	379
and all this sort of thing	4	0	0
and all that sort of stuff	3	0	0
Total	1537	1770	971

These results point to a broader range of forms being used in British casual conversation data and, to a lesser degree, in Irish English conversations whereas only seven forms were identified in the academic data. From the quantitative results, we also see that academic discourse seems to make less use of VCMs than casual conversation. We also see that one form in particular, *and so on*, in the LIBEL data, accounts for almost 40% of all VCMs used. An extract from the concordance lines of this form illustrate that it is used with a broad range of exemplars from different disciplines.

Figure 1 – Concordance samples of *and so on*

you know **the Irish having a pig in the parlour** and so on.
you have like a **play button or a stop and a rewind next and forward** and so on.
to exert relative to **the the actual height of the workstation** and so on.
from an ergonomic viewpoint in relation to ahh **the amount of force** and so on.
we have **several different lists of ahh guidance for workplace design** and so on.
it can also contribute to **the accidents** and so on.
The child itself gets its better at amm **going to different people** and so on.
Well the bottom line is you will have people who are both **tall and short** and so on.
So amm so dexterity it's **your ability to be able to manipulate objects** and so on.
So actually the average tax rate could be **twelve thousand to zero** and so on.
You have to have **fresh blood going into various muscles** and so on.
This file again by the twenty five hours. This by the **forty five hours** and so on.
Amm and we can also then consider the actual **workplace height** and so on.
one cry might mean lion yeah. Another cry might mean **danger** yeah and so on.

The narrower range of forms found in the LIBEL data also offers us useful information as to lexical choice in the more formal context of academic discourse. The forms which prevail (*et cetera, or something like that, and so forth (like that), (and) that kind of thing, and so on and so forth and and so on*), apart from *or something like that*, are either not used in casual conversation or are used to a much lesser degree. *Or something like that* seems to have a wide and unmarked usage in all three datasets.

In terms of finding out more about why speakers use VCMs in academic discourse, we turn now to a qualitative analysis which uses the four modes as its framework.

Managerial mode

In extract 3 below, we are at the beginning of a small group seminar on oral history, with the lecturer setting up an activity and organizing the seating so that the session can begin. In this extract, as in others where managerial mode is prevalent, there is no evidence at all of vague language. Instead, the lecturer makes extensive use of instructional language (*if you do have access to one of those transcripts; Just make sure you sit beside someone*

you can look in with) to locate the teaching and learning in time and space (*all you do is pull the chair over by somebody who has one; I know a lot of people weren't here last week for very good reasons ah just all you can do is fill in whatever words of wisdom were spread around ah from other people's notes*). Managerial mode occurs most often at the beginning of a piece of teaching and is characterized in the first instance by an extended teacher turn of more than one clause and a complete absence of learner turns. The focus is on the 'institutional business' of the moment, the core activity. Typically, there is a considerable amount of repetition and some kind of 'handing over' to the learners which occurs at the end of each sequence. At this point, there is a movement to another mode: in extract 3, for example, the pedagogic focus is re-aligned away from directing learning (managerial mode) to analysing a tape script (skills and systems mode).

Extract 3

<\$1> It's an awful setting in the way the room is at the moment but aam if I try and around a few and all you do is pull the chair over by somebody who has one. Aah Yeah okay hopefully. Ah I'd like to make sure now about the tape and the volume is the volume is there. Yeah you might need to bring it up. Anyway look right folks we'll start. Ok it's very awkward. It's not the kind of set up we'd like to have because the lines are too reminiscent of what's going to happen in a week or two but it's not very pretty but anyway sure we'll do the best we can. Now aam I know a lot of people weren't here last week for very good reasons ah just all you can do is fill in whatever words of wisdom were spread around ah from other people's notes aam and if you do have access to one of those transcripts eh all the better. Just make sure you sit beside someone you can look in with.

Where vague language does occur, it appears to function almost as a time-saving device so that the main item on the teaching agenda can be realised with minimal disruption and minimal waste of time. Compare extract 4 below from a different lesson in another discipline. Here, the lecturer is anxious to move on to the task and to engage students

with their own data which they were required to collect as part of their assignment for the semester, as part of a media class.

The VCM *and so on* serves to minimise the time spent on setting up the task and allows the teacher to ‘hand over’ to learners with minimal fuss. The vague category *audience agency and so on* is taken as a given, something that they already know about from recent input. The VCM here stands to mark shared/given knowledge which is background to the task at hand.

Extract 4

Really what I want to know when you having done the interview and scribed it and looked at the content of the interview how does it relate to how you understand audiences and you now understand more about audiences, about audience agency **and so on**. Then you will obviously feed into your concluding points about the particular interview, about how it went, about what the content of the interview has taught you in terms of audience based research.

Here is another similar example from a physiotherapy lecture where the VCM is found in the context of setting the scene for the next stage.

Extract 5

So there may be some accidents and maybe some injuries and maybe some a strong physiological stress on the body. Especially if maybe it’s ahh a hot environment or a very cold. Once again you can use some subjective assessments to actually to assess how the person ask the person if they’re fatigued in the course of this task **if a body was lumbered and ahh things like this**. That’s just kind of setting the scene. And we will be coming across some ahh more points like for example Corlett’s principles in the next ahh few lectures. I’m going to use some specific points which we will consider here in relation to the machine design and the operator so that we can reduce the problems for

example with repetitive strain style injuries. Okay so. Next we're going to have a look at amm evaluating the solution...

To sum up then, we can say that there is little evidence of vague language in managerial mode owing to lecturers' concern to establish a meaningful context where learning can take place. Any examples which do emerge in the data only serve to facilitate the process of setting up (or feeding back on) an activity, or organizing learning in the most effective way. Throughout, the prime pedagogic goal is to transmit information in the most economical way. Being able to use a VCM to refer to assumed background knowledge at the start up phase helps expeditious value for the lecturer.

Materials mode

Materials mode centres around a phase in a lesson where there is input or practice around a piece of material, responses are elicited in relation it and concepts and comprehension is checked. This mode is not one in which we find VCMs. As extract 6 illustrates, interaction within this mode comprises many short IRF exchanges. The language is very specific and vague language, of any type, is rare here.

Extract 6 [<\$1> = lecturer]

<\$1> Now let's correct our homework from amm Tuesday night. Tuesday's homework. Was everybody here? For this? No? Chad maybe you weren't here were you? He has it yeah. Yeah. Perfect perfect. Yeah yeah yeah yeah yeah yeah yeah. Okay. So amm A ahh Renato?

<\$3> Ahh Ahh by.

<\$1> By. Okay very good. Can you read it for me please?

<\$3> <\$G?>.

<\$1> Okay very good. Ahh lee. <\$G?>.Ley Ley ley there. Ley.

<\$4> He is seven.

<\$1> Very good. Can you read it for me there please.

<\$4> <\$G?>.

<\$1> No no the question.

<\$4> Oh read the question.

<\$1> Okay.

Five turns later

<\$1> And decorate the? The? Decorate the?

<\$2> Flat.

<\$1> Flat flat flat. Not a flight.

<\$2> Flat.

<\$1> Okay. Pronounce that word.

Skills and systems mode

In this mode, the interaction evolves around the core subject of the particular discipline. The main pedagogic goals are to allow students an opportunity to familiarize themselves with new skills or concepts and to provide corrective feedback. The discourse is typically tightly controlled and teachers make frequent use of display questions to elicit responses which are then evaluated. Meanings may be clarified in the give-and-take of the interaction through error correction, requests for clarification and confirmation checks.

Where new concepts are expressed by technical language, teachers may *scaffold* key terminology, offering learners an opportunity to gain access to a discourse community through the language of that community. Vague language does occur in this mode, as illustrated in extract 7 below.

In the data, the lecturer is under time pressure and uses vague language (*and so on*) as a means of reducing his contribution. As he recaps, he avoids the need to re-list the points which have been covered in the earlier part of the lecture, allowing students an opportunity to recall that information for themselves (*rejuvenating the economy, poverty alleviation and so on*). But, and perhaps more importantly, the vague language expressed in *and so on* does more than save time and prompt students to recall what has been covered earlier in the lecture. This example of vague language also creates a sense of shared space, common ground. The lecturer here, through his use of vague language is actually saying ‘we all know this - I don’t need to repeat it for you’. The net effect of this is to ensure that learners feel included and feel ‘safe’ as opposed to feeling intimidated or excluded.

Extract 7

Lecturer: Okay and again equity $\langle G \rangle$ so I’ve mentioned there rejuvenating the economy poverty alleviation **and so on**. They’re the kinds of equities that we’re the effectiveness of the redistribution of taxation $\langle G \rangle$ amm Okay **and so on**. Now we can only cover this to a certain degree. We’re very limited by the amount of time.

Shared space and inclusive language are crucial to successful teaching since they create an atmosphere in which learners are prepared to take risks and offer their own perspective on the content of the lecture or seminar. In extract 8, learners are made to feel included

and this is part of the process of collaborative meaning-making which is so important in higher education discourse. Here, the lecturer is giving students an opportunity to answer without making them feel trapped or intimidated by the question.

Extract 8 [<\$1> = lecturer, <\$2> = student]

<\$1> ... did you put it on V H S then or or ah

<\$2> yeah

<\$1> excellent did you try and digitize it or put it on the web or **anything like that**

<\$2> totally <\$G>

<\$1> oh very good excellent excellent

The VCM *anything like that* offers options to the student and also creates shared space in which students feel free to respond. A more direct question such as ‘did you put it on the web’ might have been interpreted as a criticism and not received any response from students – the phrase *anything like that* functions as a ‘softener’, oiling the wheels of the interaction, making the question less direct and facilitating a sense of ownership.

In extract 9, we see that the use of a VCM by a student allows for the tentative positing of an answer to the lecturer’s question. This hedging effect of the VCM here provides face protection for the learner as well as marking the proposition as tentative.

Extract 9 [<\$4> = lecturer, <\$7> = student]

<\$4> Okay. In amm nineteen eighty eight and nineteen ninety one there was a labour force survey done in each year. Now I’m just going to show you what sectors that ahh they were concerned with. Okay? Now how about someone anyone hazard a guess. Just analyse the graph analyse the bar graph now. Why do you think agriculture is so low and services is so high? Mike?

<\$7> I don’t know agriculture. You know fixed pay **and things like that**. ... More people going to college more people coming out of college. Better jobs going there

We also note the use of the pragmatic marker *you know* in conjunction with the VCM. As noted by Carter and McCarthy (2006) *you know* projects the assumption that knowledge is shared or that assertions are uncontroversial, and reinforces common points of reference. The use of *you know* plus the VCM *and things like that* serve to tentatively project shared knowledge on the part of the student. Jucker, Smith and Lüdge (2003) point out that vague category construction asks the hearer to construct the relevant components of the set which they evoke and, in so doing, promote the active cooperation of the listener. In the learning context of the LIBEL data, we could say that VCMs are also a vehicle for learning. When they are used, in skills and systems mode, on the part of the lecturer, they promote active cooperation that results in learning. When they are used by learners, they also engage cooperative peer-to-peer engagement with the category and reach out to the teacher for confirmation. They therefore provide evidence of learning in action.

Classroom context mode

In classroom context mode, the management of turns and topics is determined by the local context; opportunities for genuine communication are frequent and the teacher plays a less prominent role, allowing learners all the interactional space they need. The principal role of the teacher is to listen and support the interaction, which frequently takes on the appearance of a naturally occurring conversation. Pedagogic goals typically centre on promoting dialogue and discussion; students have genuine opportunities to express their own ideas and to make real contributions to academic debate. Learner responses are usually quite long and the teacher may offer scaffolded input or seek clarification as and

when it is needed. Vague language functions here in much the same way as it functions in everyday conversation; for example, it acts as an ‘involvement device’ ensuring listener participation and promoting equity and understanding.

Consider extract 10 below. Here the teacher is trying to make a point by using a literary reference. The whole style and register of this extract suggest that it could just as easily have been taken from naturally occurring conversation. The VCMs here (*and stuff, and stuff like that*) ensure that the listeners feel involved and that there is empathy towards the stance that the teacher adopts. In this way, the teacher is able to progress the discourse, bringing everyone along together and making sure that there is a sense of purpose and direction to the dialogue. Again, the vague language being used here serves to ‘soften the blow’ of a more didactic tone. A more conversational style is also almost certain to promote good listenership (McCarthy 2002, 2003) and means that the learning will be more memorable.

Extract 10

Lecturer: ...did any of you ever read Angela's Ashes? Yeah exactly and it's just it's just the poems **and stuff** that the Daddy keeps on you know every time he has a few drinks and he's living abroad and he's broke and he's after like leaving Ireland like arrived there filled with the pox and you know like. It was just like not at all a romantic story. He gets there and then before you know he's like standing up all the kids at night time going we'll die for Ireland. And you know there's was all of these like poems and **and stuff like that** and it was all about like will you die for Ireland?

Classroom context mode, then, offers the greatest potential for vague language since it most closely resembles naturally occurring conversation. Note that in this mode, vague language is as likely to be used by students as it is by teachers, as exemplified in extract

11. Here, the student asks a question, but uses vague language (*and everything*) as a means of creating shared space and involving the teacher-listener (<\$1>). The net effect of this is to promote understanding and to ensure that the questioner is fully understood.

Extract 11 [<\$6> = student, <\$1> lecturer]

<\$6> I have a question.

<\$1> Yeah?

<\$6> I was reading in one of our books that ethnicity and race are completely different things and ethnicity you learn things and race is a ahh is inherent in the you know in the blood and your appearance **and everything**. That is that wrong?

<\$1> It depends on what theorist you go after.

Conclusions

To summarise, while the corpus findings point to fewer VCM forms and lower occurrences of VCMs in the academic data compared with casual conversation, we find that they have some specific uses within this academic discourse. These functions tie in with pedagogical goals of the interactional mode within which they occur. Within *managerial mode*, VCMs can be used by the teacher to help expedite the start up phase of a lesson or activity. Because they provide shortcuts that mark information or concepts that can be taken as given, shared or unproblematic, they very quickly set up what is common ground and facilitate a speedy handing over to the task phase of the lesson. In *skills and systems mode*, they operate as two-way portals. For the teacher, they can open a door to what is key shared knowledge for this phase of the lesson and create a shared space around this commonage. For the learner, they open a door to a space where it is safe to take risks. Tentative propositions can be marked using VCMs and loss of face is avoided. In this mode, they engage cooperative listenership on the part of peers which also facilitates learning. Finally, in *classroom context mode*, we find that language use, in general, more resembles casual conversation (see Walsh 2006) and here VCMs occur along with other vague language items and mark shared uncontested knowledge.

From a second language perspective, it is clear that the ability to understand and create VCMs is an important part of classroom language. Their prevalence in terms of high frequency chunks in casual conversation also adds to the case for including them as vocabulary items in English for Academic Purposes programmes.

At a methodological level, we have used our quantitative corpus findings and comparisons as a starting point and delved into the data in a qualitative way. The modes framework provides a useful lens for looking at academic data and there is great potential for synergies with corpus linguistics.

References

- Barsalou, W.L. 1983. Ad hoc categories. *Memory and Cognition* 11: 211-77.
- Barsalou, W.L. 1987. The instability of graded structure: implications for the nature of concepts. In *Concepts and Conceptual Development*, U. Neisser, (ed), 101-40, Cambridge: Cambridge University Press.
- Carter, R. A. and McCarthy, M. J. 1995. Grammar and the spoken language. *Applied Linguistics* 16 (2): 141-58.
- Carter, R. A. and McCarthy, M. J. 2006. *Cambridge Grammar of English: A Comprehensive Guide to Spoken and Written English Grammar and Usage*. Cambridge: Cambridge University Press.
- Chafe, W. 1982. Integration and involvement in speaking, writing, and oral literature. In *Spoken and written language: exploring orality and literacy*, D. Tannen (ed.), 35-53. Norwood, N.J.: Ablex Publishing Corporation.
- Channell, J. 1994. *Vague language*. Oxford: Oxford University Press.
- Cruse, D.A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press
- Cutting, J. 2000. *Analysing the Language of Discourse Communities*. Oxford: Elsevier Science.
- Dines, E. 1980 Variation in discourse – and stuff like that. *Language in Society* 1: 13-31.
- DuBois, S. 1993. Extension particles, etc. *Language Variation and Change* 4: 179-203.

- Evison, J., McCarthy, M.J. and O'Keeffe A. (In press) *Looking out for love and all the rest of it: vague category markers as shared social space*. In *Vague Language Explored*, J. Cutting (ed.). Basingstoke: Palgrave.
- Franken, N. 1997. Vagueness and approximation in relevance theory. *Journal of Pragmatics* 28: 135-151.
- Heritage, J. 1997. Conversational Analysis and Institutional Talk: Analysing Data. In *Qualitative Research: Theory, Method and Practice*. D. Silverman (ed.). London: Sage Publications.
- Heritage, J. and Greatbatch, D. 1991. On the character of institutional talk: The case of news interviews. In *Talk and Social Structure*, Boden, D. and Zimmerman, D. (eds.), 359-417. Cambridge: Polity Press.
- Jarvis, J. and Robinson, M. 1997. Analysing Educational Discourse: an Exploratory Study of Teacher Response and Support to Pupils' Learning. *Applied Linguistics* 18 (2): 212-228.
- Jefferson, G. 1990. List construction as a task and resource. In *Interaction Competence*. Psathas, G. (ed.). Lanham, MD: University Press of America, 63-92.
- Jucker, A.H. Smith, S.W. and Ludge, T. 2003. Interactive aspects of vagueness in conversation. *Journal of Pragmatics* 35: 1737-69.
- Kumaravadivelu, B. 1999. Critical Classroom Discourse Analysis. *TESOL Quarterly* 33(3): 453-484.
- Macaulay, R.K.S. 1991. *Locating Dialect in Discourse: the language of honest men and bonnie lasses in Ayr*. New York: Oxford University Press.
- McCarthy, M. J. 1998. *Spoken Language and Applied Linguistics*. Cambridge: Cambridge University Press.
- McCarthy, M. J. 2002. Good listenership made plain: British and American non-minimal response tokens in everyday conversation. In *Using Corpora to Explore Linguistics Variation*. R. Reppen, S. Fitzmaurice, and D. Biber (eds.), 49-71. Amsterdam: John Benjamins.
- McCarthy, M. J. 2003. Talking back: 'small' interactional response tokens in everyday conversation. In Special issue of *Research on Language and Social Interaction* on 'Small Talk'. J. Coupland (ed.) 36 (1): 33-63.

- McCarthy, M.J. and Carter, R.A. 1995. Spoken Grammar: what is it and how do we teach it?. *ELT Journal* 49(3): 207-218.
- McCarthy, M.J. and Carter, R.A. 2002. From conversation to corpus. In *Windows on the World: Media Discourse in English*, A. Sánchez-Macarro (ed.), 15–39. Valencia: University of Valencia Press.
- Mervis, C.B. and Rosch, E. 1981. Categorization of natural objects. *Annual Review of Psychology*. 32: 89-115.
- O’Keefe, A. 2003. ‘Like the wise virgins and all that jazz’ – using a corpus to examine vague language and shared knowledge. In *Applied Corpus Linguistics: a multidimensional perspective*, U. Connor and T. A. Upton (eds.), 1-20. Amsterdam: Rodopi.
- O’Keefe, A. 2006. *Investigating Media Discourse*. London: Routledge.
- Overstreet, M. and Yule, G. 1997a. On Being Explicit and Stuff in Contemporary American English. *Journal of English Linguistics* 25(3): 250-58.
- Overstreet, M. and Yule, G. 1997b. Locally Contingent Categorization in Discourse’, *Discourse Processes* 23: 83-97.
- Pavlenko, A. and Lantolf, J.P. 2000. Second language learning as participation and the (re)construction of selves. In *Sociocultural Theory and Second Language Learning*, J.P. Lantolf (ed.), 155-178. Oxford: Oxford University Press.
- Rosch, E. 1978. Principles of categorization. In *Cognition and categorization*, E. Rosch and B. Llyod (eds.), 27-48. Hilldale, N.J.: Erlbaum.
- Rosch, E., C.B. Mervis, W.D. Gray, D.M. Johnson and P. Boynes-Braem 1976. Basic objects in natural categories. *Cognitive psychology* 2: 491-502.
- Scott, M. 1999. *Wordsmith Tools Software*. Oxford: Oxford University Press.
- Seedhouse, P. 2004. *The Interactional Architecture of the Second Language Classroom: a Conversational Analysis Perspective*. Oxford: Blackwell.
- van Lier, L. 1988. *The Classroom and the Language Learner*. London: Longman.
- Walsh, S. and O’Keefe, A. (In press) Applying CA to a modes analysis of third-level spoken academic discourse. In *Conversation Analysis and Languages for Specific Purposes*, P. Bowles, and P. Seedhouse (eds). Peter Lang
- Walsh, S. 2006. *Investigating Classroom Discourse*. London: Routledge.

- Ward, G. and Birner, B. 1992. 'The semantics and pragmatics of "and everything"', *Journal of pragmatics*, 19: 205-214.
- Zhang, Q. 1998. Fuzziness – vagueness – generality – ambiguity. *Journal of pragmatic*, 29: 13-31.

Appendix 1 – Transcription conventions

<\$X> speaker turn, e.g. <\$1> = speaker 1, <\$2> = speaker 2 etc. in order of 'appearance' on the recording.

<\$G?> ... <\$G2> <\$G?> marks uncertain or unintelligible utterances where the number of syllables cannot be guessed. Where the number of syllables can be discerned, this number is marked, e.g. <\$G2> denotes two intelligible syllables.

¹ Hereafter, LIBEL CASE will be shorted to LIBEL.

² At the time of writing, LIBEL comprises one million words, 500,000 of which is fully transcribed.

³ Note that while Limerick and Belfast are geographically on the same island (of Ireland), they come under two different jurisdictions: 1) The Irish Republic and 2) The United Kingdom and Northern Ireland, respectively.