CHAPTER 8

'LOOKING OUT FOR LOVE AND ALL THE REST OF IT':

VAGUE CATEGORY MARKERS AS SHARED SOCIAL

SPACE

Jane Evison and Michael McCarthy

University of Nottingham, UK

Anne O'Keeffe

University of Limerick, Ireland

Introduction

Carter and McCarthy (2006, p.202) assert that VL expressions are a strong indication of an assumed shared knowledge and that they mark 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 VL as a

marker of in-group membership. It is this interactive aspect of VL that we will focus on in this chapter. We examine one particular manifestation of vagueness: the creation of vague category markers (hereafter VCMs), such as 'university courses and that sort of thing'; 'I've got to wash my hair and everything', where speakers refer obliquely to other members of categories which they assume their listeners will be able to 'fill in'. In extract (1) from an everyday conversation at a family dinner table (taken from the Limerick Corpus of Irish English, hereafter LCIE) where the participants are talking about someone who has taken a job at a local fast-food restaurant, one of the speakers throws out an ad hoc category (Barsalou 1983, 1987):

(1)

Speaker 1: And what's he going to be doing in there?

Speaker 2: I think they're training him as a trainee manager.

Speaker 1: Frying chips?

Speaker 3: You mean he's frying chips. Basically. <laughs>

Speaker 2: He says 'I'm going to do everything. Fry chips and wait tables
         and stuff'.

Speaker 1: …there's no way he'll be able for that like <laughs>

The category that speaker 2 creates did not derive from any pre-fabricated lexical chunk before he spoke 'fry chips and wait tables and stuff'. Yet the speaker needed this category in this situation and he had it within his resources to create it. He did so in the knowledge that his interlocutors would know what it meant and cognitively that they would be able to fill in the set that he has referred to within their shared cultural frame of

reference. In the (Irish) context in which the category was projected the set refers to the range of possible activities that one could be asked to undertake while working in a fast-food outlet, such as cook chips and burgers, serve customers, sweep floors, clean tables, but not paint walls, design advertising, do book-keeping or sing to the customers. The set has a finite range, the limits of which are understood within the socio-cultural context, and for the speaker to have listed every possible item in the set would have been at best pedantic and at worst absurd. Jucker, Smith and Lüdge (2003) point out that such vague categories ask the hearer to construct the relevant components of the set which they evoke and promote the active cooperation of the listener. Some more examples of the VCMs under scrutiny in this chapter are given here (taken from the CANSOC corpus; see below).

(2)

[Speaker is talking about various people's jobs]

And my husband travelled for his father, selling and that sort of thing.

(3)

Speaker 1: He was interested in keeping bees.

Speaker 2: Oh yes, yes, bees and chickens and all the rest of it.

(4)

She frames pictures and so on and she doesn't have much free time.

The rationale behind this chapter is that in order to use VCMs successfully, speakers must have expectations about what their co-participants know, and that such expectations are negotiated within social space, in the sense expounded by Vygotsky

(1978), for whom social relationships, language use, thought and cultural activity share the same creative space (see section 2 below). Within a socially defined group, VCMs become a tool for creating short-cuts.

In this paper we use three spoken corpora, two sub-corpora of the predominantly British English CANCODE[1] corpus, and the Limerick Corpus of Irish English (LCIE), to explore VCMs in contexts where the participants have different degrees of shared knowledge and intimacy. We explore how the shared knowledge required on the part of the participants in order to interpret VCMs has a common core of socio-culturally ratified 'understandings' in each specific context and that the range of domains of reference of these categories is relative to the assumed depth of shared knowledge of the participants and their social relationships.

VCMs are most typically, but not exclusively, found in clause-final positions and often consist of a conjunction and a noun phrase (for example, 'and/or that sort of thing'). In the literature, they 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). In this chapter we adhere to O'Keeffe's terminology.

O'Keeffe (2003) refers to VCMs as recognisable chunks of language that function in an expedient way as linguistic triggers employed by speakers and decoded by co-participants who draw on their store of shared knowledge. In a corpus-based study of an Irish radio phone-in (whose data is called upon in the present chapter) O'Keeffe argues that the meanings of vague categories are socio-culturally grounded and are co-

constructed within a social group that has a shared socio-historic reality. This is consistent with Overstreet and Yule (1997b), who point out that the process of establishing categories is locally contingent in discourse.

In another corpus-based study McCarthy, O'Keeffe and Walsh (2005) compared VCMs and their referents in three corpora, the five-million-word CANCODE corpus, LCIE, a one-million word corpus of Irish casual conversation, and the Limerick and Belfast Corpus of Spoken Academic Discourse (LIBEL), a one-million word corpus of academic discourse collected on the island of Ireland. They noted differences in VCMs in the academic data compared with the casual conversation corpora (LCIE and CANCODE). For example, 'et cetera', widely used in the academic context, was rare in the conversational ones; additionally, the academic context showed VCMs functioning to hedge factual assertions more than in conversation. O'Keeffe (2006) further compared these findings with a sample of VCMs from a corpus of media discourse and found that the forms used in political interviews most resembled those in the academic discourse from LIBEL. She also noted, like McCarthy, O'Keeffe and Walsh (2005), that the more institutionalised data contained fewer instances of vague categories. McCarthy, O'Keeffe and Walsh (2005) also found the participants in a university small group setting drew on shared knowledge and, influenced by the work of Vygotskian applied linguists (see also section 2 below), they suggested that vague categorisation was a means of the creation and maintenance of 'shared space' within this classroom setting, and a significant site for learning opportunities and concept-formation.

Such studies seem to point to the use of VCMs as purposeful, creative and highly interactive. In this chapter we hope to reinforce those views and to examine in greater detail how speakers in different contextual domains make reference to collective

phenomena and experiences in ways which their interlocutors can decode and share within particular contexts, and thus enter into that social space where language and thought co-exist and push into new conceptual frontiers.

A Vygotskian Perspective

VCMs are, above all, highly interactive: they invite the interlocutor to enter a conceptual space with the speaker where phenomena perceived as sharing characteristics are bundled together in acts of meaning-making. Those phenomena are 'projected' as shared experience; there is never any guarantee that two or more minds are conceptualizing the full range of identical phenomena. Such creative activity within the shared space enables new acts of cognition, whether these are instrumental in crystalising new stances, opinions, judgements or simply different personal perspectives on people and things in the social and cultural environment. Of relevance here is Vygotsky's notion that social interaction plays a fundamental role in the development of cognition. Speaking of child development, Vygotsky (1978, p.57) asserted:

> Every function in the child's cultural development appears twice: first, on the social level, and later, on the individual level; first, between people (interpsychological) and then inside the child (intrapsychological). This applies equally to voluntary attention, to logical memory, and to the formation of concepts. All the higher functions originate as actual relationships between individuals.

Much of what Vygotsky says concerning the child's social experience is relevant to our present concerns. Even more relevant is the child's proclivity to categorise. It is an uncontroversial observation that a child naturally sorts things into categories which share common attributes, a process which, in its initial stages may produce categories which the adult perception would dismiss (for example, calling a sheep a 'dog' because it has four legs and a fluffy coat). Such attempts at basic categorization Vygotsky refers to as 'diffuse complexes' (1962, chapter 4). Diffuse complexes enable generalities to be made based on concrete experience by perceiving similarities among phenomena, however unstable such perceived similarities may be.

VCMs capture the fluidity and instability of the diffuse complex, the pre-conceptual phase where the language user attempts to make meaning from diverse phenomena and experiences, and reaches out to his/her interlocutor in an appeal to equally diffuse and unstable shared experience. Within the social space of such negotiations, it is not just language which is creative, but thought itself, and the language user has the possibility of new understandings and new critical, ethical and moral positions (Crawford 2001), whether in the pedagogical context of the school or university, in the public media context of broadcast debates, radio phone-ins, and so on, or in the private and intimate fora of casual conversation. It should not be a source of surprise, therefore, that what in the lay perception may be typically characterized as sharp, focused discourse and incisive intellectual exchange (for instance, academic discussion or broadcast debate) should in fact be frequently characterized by the same kinds of vague references to non-institutionalised and only partially formulated categories of external phenomena and human experience as occur in casual conversation, as we hope to show in this chapter. Vague categories are far from vague in

the negative sense of uninformative or sloppily constructed; they are at the creative forefront of language use and the collaborative making of meaning.

Data And Methodology: CANCSOC As Benchmark

This chapter bases its initial, wide-ranging analyses on a one-million-word sub-corpus of the CANCODE spoken corpus. The sub-corpus consists only of a sample of socialising and intimate conversations, and excludes professional (such as workplace conversations), transactional (for example, service encounters) and pedagogical (for instance, the university classroom) conversations. This last group are addressed separately in this chapter (the CANCAD corpus, see below). We refer to this socialising sub-corpus as CANCSOC.

The investigation began with an analysis of 'chunks' in the CANCSOC corpus. The analytical software used (*Wordsmith Tools*; Scott 1999) is capable of automatically retrieving recurrent strings of words and generating frequency lists for their occurrence. In this chapter we focus on those items from the first 500 (or the whole list where this is less than 500; see below) of the automatically generated rank-order frequency lists for CANCSOC which display the potential to act as VCMs.

Rank-order frequency lists of two-, three-, four-, five- and six-word sequences were generated. The lists were then combed for all items occurring ten times or more which could potentially act as VCMs. These were then checked against concordances to see if they were in fact used in this way. For the longer lists (the two-, three- and four-word ones), only the first 500 items were considered, and only items which formed

complete 'chunks' (that is to say, which displayed syntactic and semantic/pragmatic integrity, see below) were extracted. The resultant VCM chunks are presented in rank order of frequency in Table 8.1.

Table 8.1   VCMs in CANCSOC

_____

Round brackets indicate lexical items that may co-occur.

Items within square brackets are alternative but mutually exclusive (for instance, 'that [kind/sort/type]' of X implies 'that kind or sort or type of X').

VCMs	Total
and/or [something/anything/everything] (like that)	1024
(and/or) (X) stuff (like that/X)	620
and (all) (of) that	270
(and/or) thing(s) (like that/X)	579
(all) [this/that/these/those] [kind(s)/sort(s)/type(s)] of X	219
(or) whatever	90
and so on (and so forth)	60
et cetera (et cetera)	30
Xs like that	25
and all the rest of it	12
(and) this that and the other	11
Total	2940

Many recurrent strings, although frequent, do not qualify as potential VCM chunks as they do not display syntactic or semantic/pragmatic integrity, for example, 'that sort of', 'and stuff like'. These are often incomplete segments of longer strings which do possess wholeness ('that sort of thing', 'and stuff like that'). However, some strings can be both whole in themselves and form part of longer strings, for instance, 'and that', which functions as a VCM in CANCSOC (for example, 'The fans you get in Spain are all these fancy ones with lace <u>and that</u>'), but which is also part of the longer chunk '<u>and that</u> [kind/sort] of thing'. The total frequency counts were therefore performed by subtracting and listing separately the totals for shorter, integrated items where they also occurred as part of longer items.

Items were demarked according to their syntactic headwords: for example, items with 'kind/sort/type' as headword ('all these <u>kinds</u> of things', 'that <u>sort</u> of thing') were listed separately from items with 'thing(s)' as headword ('<u>things</u> like that', 'and <u>things</u>'). The limited scope of the CANCSOC count, focusing only on high-frequency items, does not take into account items which may operate as VCMs but which are simply not sufficiently evidenced. For example, one utterance which clearly contains a VCM is 'I was sitting with Jim <u>and that lot</u>', where a high degree of shared knowledge is presupposed (who the members of Jim's surrounding group were). However, 'and that lot' only occurs as a VCM seven times, falling below the CANCSOC cut-off point of ten occurrences. This issue is even more acute in the case of the two smaller, specialized corpora used for comparison (CANCAD and *Liveline*; see below). For those corpora, the CANCSOC VCMs were checked and, in addition, the two corpora were read line by line and all other VCMs, even those occurring only once, were manually added. These

manually added VCMS were then back-checked in CANCSOC and any occurrences were added to the CANCSOC total.

CANCSOC And Types Of VCM Reference

CANCSOC is a corpus of informal conversations among friends and intimates, so it is not surprising that CANCSOC VCMs encode a high degree of projected shared knowledge, often knowledge which is shared widely within the British and Irish speech communities. This means that in many cases, any member of those communities (or indeed people beyond the communities) can successfully 'fill in' the category members. However, in many cases, the categories are opaque, to the extent that category members are obscure or can only be speculated upon by the non-participant observer-analyst. A range of examples ranked from transparent to opaque serve to illustrate this (Table 8.2).

The VCM examples in Table 8.2 are all based round noun phrases, but VCMs may also refer to categories of states, actions and events:

(5)   I wasn't expecting to be sort of <u>judged and criticised and things.</u>

(6)   It was really good, it was <u>sunny and everything</u>, not at all cold.

(7)   I'll be super fit, not <u>out of breath or anything</u>.

Table 8.2   Examples of VCMs in CANCSOC

CANCSOC example	Comments
'She appreciates quietness and peace and she loves <u>flowers and that sort of thing</u>.'	easily interpretable by most people anywhere in the world
'We'll meet up and go to Leeds for the day because there's <u>a new Marks and er a new Debenhams and stuff like that</u>.'	less easily interpretable; one needs to know that *Marks* and *Debenhams* are large departments store chains; most British/Irish people know this
'So like God speaks to us through the prophets as well now doesn't he. <u>Likes of Tony Ling and that</u>.'	interpretable only by those people with knowledge of *Tony Ling* as a religious figure within a minority Christian sect
[speech at a family birthday party] 'Four generations here today and that's important. And Mrs Wheeler and my dad are of the first generation. And then there's old ones like <u>Bobby and Paul and so on</u>.'	only interpretable by the family members and others at the social gathering who know the family

Two Comparative Corpora: *Liveline* And CANCAD

We now turn to look at smaller amounts of contextually situated data. The results from CANCSOC will form a baseline against which these other data can be compared. Here we use two small corpora. The first is a 55,000 word sub-corpus of LCIE, consisting of data from an Irish radio phone-in show called *Liveline* which is broadcast every weekday on *Radio Telefís Éireann* (RTÉ). The programme has been running for 20 years and has an audience of over 365,000[2], almost 10 per cent of the Irish population. These data were taken from a sample of programmes in 1998, comprising 44 phone calls from a total of five programmes. Topics of calls to the show include the following miscellany: female facial hair problems, tattoos, the peace process in Northern Ireland, how ears were pierced in the old days, warnings about the decline of fidelity and moral decay in general, and the growing trend of litigation in Irish society, among others (see O'Keeffe 2003).

The second small corpus, CANCAD (Cambridge and Nottingham Corpus of Academic English) is composed of seven university seminars taken from the pedagogical section of the five-million-word CANCODE corpus. There were a range of speaker styles and approaches evident in the seminars: in two the tutor held the floor for a considerable portion of the time, in four the tutor led a whole-class discussion throughout the session, and in one the tutor left the room during the discussion. In four out of the seven seminars, participants were ready to talk about literature texts they had prepared for the class and in the other three handouts were given out at the start of the class containing textual extracts.

Liveline And CANCAD: Analysis

Liveline and CANCAD were searched for all the VCMs found in CANCSOC; this search yielded the results shown in Table 8.3.

Table 8.3   VCMs in *Liveline* and CANCAD

VCMs	CANCAD	*Liveline*
(all) [this/that/these/those] [kind(s)/sort(s)/type(s)] of X	7	17
(and) this that and the other	0	3
(or) whatever	6	10
and (all) (of) that	2	3
and all the rest of it	0	0
and so on (and so forth)	16	12
([and/or]) [something/anything/everything] (like that)	8	10
([and/or]) (X) stuff (like [that/X])	18	7
([and/or]) thing(s) (like [that/X])	43	46
et cetera (et cetera)	2	6
Xs like [this/that]	0	21
Total	102	135

A comparison of the three datasets gives the following distribution, normalised to occurrences per million words, as can be seen in Table 8.4.

Table 8.4   VCMs in CANCSOC, *Liveline* and CANCAD

Corpus	VCMs	per million
CANCSOC	2940	2940
Liveline	135	2454
CANCAD	102	1873

Even though the CANCSOC figure is only based on the first 500 items of the longer lists it exceeds that of the two smaller corpora, though the radio data is closer to CANCSOC than CANCAD is. From these initial results, we propose that the closer the speaker relationship within the participation framework (after Goffman 1981), the greater the shared space that they can exploit. The conversations in CANCSOC involve close friends and family members and have the highest number of VCMs. The radio phone-in data, as O'Keeffe (2002, 2003, 2006) has argued, involves the creation of a pseudo-intimacy within a stable participation framework. Presenters, callers and audience are attempting to create a pseudo-conversational context. Overall, *Liveline* is much more like friendly conversation than formal radio debate (see O'Keeffe 2006 for a comparison of media genres).

The academic data contains fewer VCMs than the radio phone-in data. The academic data also draws on shared knowledge, but the knowledge that is assumed within this participation framework is mostly specific to academic disciplines and academic discourse communities (Swales 1990). Swales' notion of discourse communities includes common goals and participatory mechanisms, the use of specific genres of communication, a high level of shared expertise and specialized terminology. All of these feed into the types of VCMs found in CANCAD.

In addition to the search for those VCMs of high frequency in CANCSOC, *Liveline* and CANCAD were searched manually for all occurrences of VCMs. This rendered the following additional items (Table 8.5). In order to achieve consistency, the additional VCMs were then back-checked in CANCSOC, for which the figures also appear in Table 8.5:

Table 8.5   Additional VCMs in *Liveline* and CANCAD

VCMs	*Liveline*	CANCAD	CANCSOC
Or that	2	0	2
For the X that's in it[3]	1	0	0
Or some other one of X	1	0	0
Or any of X	1	0	0
And so forth	2	1	1
Total	7	1	3

Table 8.4 now needs to be slightly adjusted to take these figures into account. The broad picture is little affected, except to bring *Liveline* even closer to CANCSOC (Table 8.6).

Table 8.6   Revised Totals for VCMs in CANCSOC, *Liveline* and CANCAD

Corpus	VCMs	per million
CANCSOC	2943	2943
Liveline	142	2582
CANCAD	103	1873

Examples of these VCMs in action include:

(8)   The doctor came down and said to her 'Oh it's just a wee bit of like diarrhoea <u>or that</u>.' (CANCSOC)

(9)   [a mother talking about her baby's symptoms of meningitis] ...his neck was sore if you had tried to move his head <u>or that</u> (*Liveline*)

(10)   [caller is complaining about lack of political debate on a constitutional amendment required as part of the Northern Ireland peace process in 1998. 'Bertie' refers to the Irish 'Taoiseach' (Prime Minister) at the time, Bertie Ahern, 'and any of them' refers to all politicians in the Republic of Ireland and Northern Ireland]

> I'd like <u>Bertie or any of them</u> get on and address what we're
>
> voting on on Friday which is the amendments to our
>
> constitutional articles two and three (*Liveline*)

As can be seen from Table 8.5, the results add very few items to the totals in Table 8.4,

and the picture remains largely unchanged (except to bring *Liveline* even closer to

CANCSOC). This suggests that the most frequent VCMs in CANCSOC are also

widespread in *Liveline* and CANCAD, though differently distributed. Both smaller

corpora, then, have features in common with banal, everyday, casual conversation.

However, as we saw in the case of CANCSOC, the category memberships signalled by

VCMs can range from universally transparent to quite opaque. We therefore now turn to

an analysis of the domains and types of references projected by the VCMs in *Liveline*

and CANCAD, in an attempt to see whether and how they reflect an appeal to the shared

space of their co-participants, how exclusive such appeals are (in terms of

interpretability by outside observers) and what their specific functions are in the contexts

in which they occur.

Reference Domains Of VCMs In *Liveline* And CANCAD

All of the VCMs in the *Liveline* and CANCAD corpora were examined in terms of their

projected referents. Broadly, the referents may be divided into 'local', 'societal' and

'global'. 'Local' is defined as interpretable by a specific group of participants and those

who share relatively exclusive social and cultural frames of knowledge, for example, a

family, a group of friends, a class of students and their teacher discussing their academic

subject. 'Societal' is defined as interpretable by all members of a speech community or

socio-political entity who share a common culture and history, for instance, English

speakers, the population of Ireland, people from a particular city or region. 'Global' is

defined as interpretable by most mature, experienced human beings throughout the

world. The results for the two corpora are shown in figure 8.1.

Figure 8.1   *Liveline* and CANCAD Reference Domains
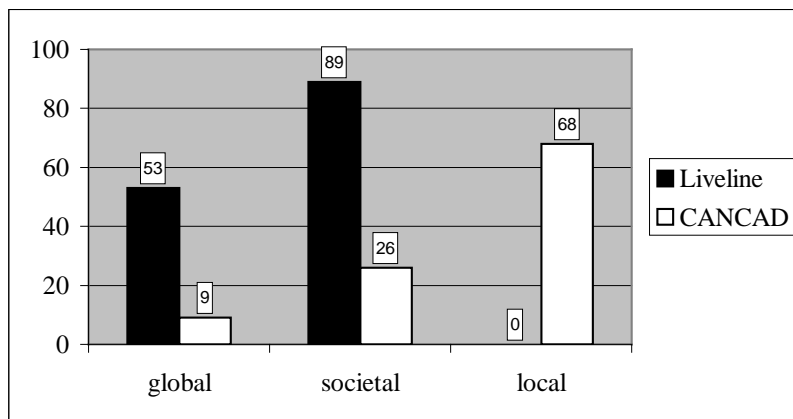
The largest shared domain of reference in the *Liveline* data was at the 'societal' level and

when this domain was further broken down, three sub-categories were identified, as

shown in Figure 8.2. When the largest of these, 'general' societal knowledge, is further

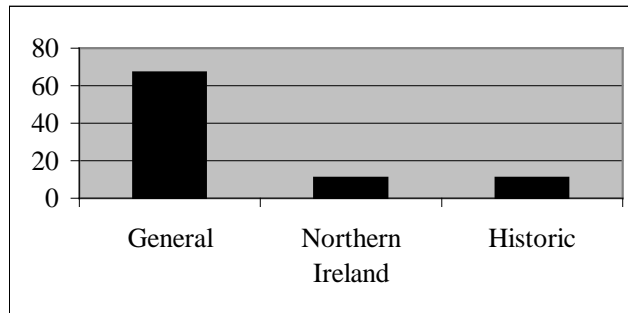broken down, we find the sub-categories listed in Table 8.7.

Figure 8.2 Societal Domains in *Liveline*

Table 8.7   Categories within the General Irish Reference Domain

Category	Example	Example utterances
Social practices and attitudes	The perception of who goes to an Irish boarding school[4]	'… of a lot of people who go to boarding schools are from unhappy families there is that kind of element…'
Social responsibi-lities and realities	Negative social realities that come with the Celtic tiger economy	'I think there's a certain anonymity about it … if we understood the real difficulties that people have in breaking out of this situations of hardship and so on'
Work and financial practices	Car rental companies in Ireland	'…car hire firms such as Dan Ryan[5], Budget, Avis people like that'
Social types	Irish criminals and social undesirables	'I think there's a lot of undesirables criminals and people like that'

The largest domain of shared space in the Irish radio data, not surprisingly, is at a general societal level. The radio phone-in callers, presenter and audience occupy this shared space and they know that they can draw on it to refer to things that will be understood, and as potential sites for new meaning-making. In contrast to CANSOC, outside observers (that is, Irish listeners to the radio programme) will not normally encounter opaque references at this general societal level. However, a listener from outside of Irish society will frequently encounter opacity, as illustrated by these examples:

Table 8.8   Examples of Societal VCS In *Liveline* Likely to be Opaque to Listeners outside of Irish Society

Example	Comment
'Didn't get <u>a Gaeltacht grant or anything like that</u>?'	A 'Gaeltacht' is an area where Gaelic is spoken. These areas get special government grants for, for example, the setting up of enterprises in the zone.
'And the Secretariat at <u>Maryfield and all that</u>?'	The Maryfield Secretariat was a joint civil service set up by the British and Irish governments under the Anglo-Irish Agreement in1985.

What one group of language users shares as its commonage is what keeps them together, like a centripetal force, while, paradoxically, this commonage can keep others away, like a centrifugal force.

The VCMs in CANCAD were classified according to reference domain and four categories which could be understood with broad societal knowledge were identified. Of these, 'language culture and gender' encompassed two thirds of all the societal VCMs, which mainly occurred as part of explanations or exemplification of points under discussion. The other three categories show more similarities with the kinds of categories of societal VCMs in the *Liveline* corpus. In this case, the VCMs tend to feature in more relational episodes, of the kind that often occur at the beginning of classes or in breaks or transitions during the class.

Table 8.9   Breakdown of Societal VCMs in CANCAD by Domain

<u>Category</u>	<u>Example</u>	<u>Example utterances</u>
Language, culture and gender	Matching items that couples have or wear	'the fact that er a woman is assumed to have a smaller car than a man <u>and so on and so forth.</u>'
Media, TV and music	Content of a film that would make it appeal to schedule writers.	'They probably saw it had some nudity in it <u>or something</u>.'
Transport and services	Evidence of poorly funded privatised railways in the UK	'The train I came across on from Birmingham to Nottingham was the most crappy train. And it it was

		marked on the outside. The seats were dirty and ripped. And the floor was dirty. <u>And everything</u>.'
University life	Activities that people expect to happen in university seminars	'Well it's just different people <u>same stuff</u>.'

By far the largest area of common reference found in the use of VCMs in the CANCAD data was at the 'local' level, that is to say, references to shared disciplinary knowledge and practices. Figure 8.3 gives a breakdown of this 'local' reference domain:

Figure 8.3   Local Reference Domains in CANCAD
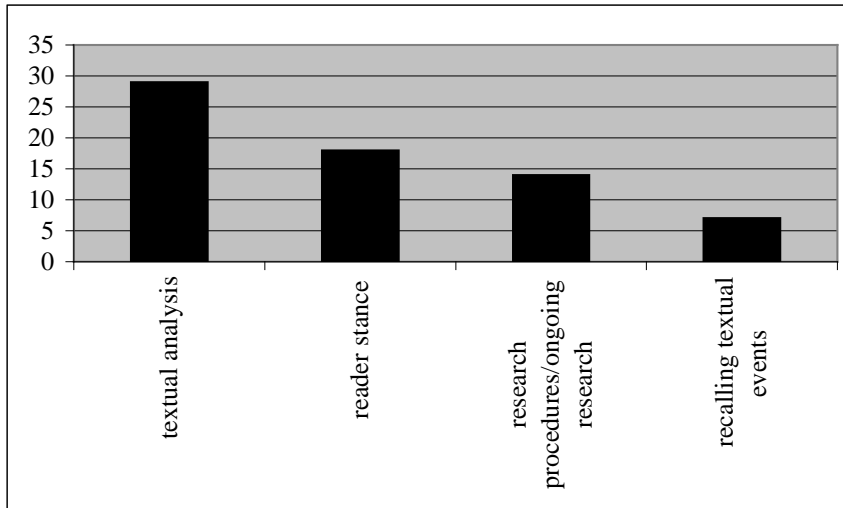
Here we see that the immediate classroom context is the locus of the greatest amount of exploitation of shared space. The classroom material, its content and interpretation, the

shared endeavour of academic activity such as understanding a text, and academic
activities such as research, appear to be where the participants of a classroom can
assume the greatest level of given and shared knowledge which can be drawn on as a
shared resource in the creation of ad hoc categories. In example (11), which occurs at
the opening stages of a postgraduate seminar on poetic language, the tutor is
encouraging the students to forget the more formal kinds of analysis they may have done
and instead react intuitively to the text.

        (11)

        Tutor: So instead of being like a machine and just thinking right I'll
                            do a discourse analysis then I'll do a pragmatic analysis then
                            I'll do a syntactic   analysis <u>and so on</u> all the way down.
                            There's no need to do all that because you can go straight for
                            your gut reaction er first time around.
                            (CANCAD)

In this next example from an undergraduate sociolinguistics seminar, the student and the
tutor are co-creating an understanding of an extract from a sociolinguistics textbook.

        (12)

        Student: A lot of insurance companies now do do <u>things like erm clean</u>
                            <u>out your car and get it fixed and stuff like that</u>. Don't they?

        Tutor: Right.

        Student: So it's not that weird

Tutor: So it's not totally weird. Yeah. It's not as weird as the police

and the washing machine.

Student: No.

Tutor: Yeah. Right. Because if it so= th= There's a like a semantic

field with things tha= <u>to do with houses burning down and</u>

<u>insuring yourself against fires and all of that.</u>

(CANCAD)

This example emphasises the collaborative nature of vague category projection. In

example (13), the students are struggling to find extracts from the novel *Mrs Dalloway*

which exemplify a list of themes given to them by their tutor.

(13)

Student 1: But I wasn't sure if I'd got the wrong end of the stick.

Student 2: Yeah. I was reading a bit in the introduction. It's like oh yeah

she was questioning her <u>sexuality and stuff.</u> Yeah.

Student 1: Yeah. I think she was.

Student 2: But then after that little bit.

Student 1: The trouble is does femin - Does femininity have anything to

do with sexuality?

Student 2: Yeah.

(CANCAD)

Student 2 refers to 'sexuality and stuff', and three turns later Student 2 questions whether this category can include femininity, a topic that she had raised earlier in the seminar, thus moving the discussion on to a new phase. Jucker, Smith and Lüdge (2003) found that explicit identification of a potential member of a vague set by another speaker helped maintain bonds between participants in casual conversation, but here in this academic discourse we can see that it also helps speakers refine academic argument and explore new conceptual territory. In the tutor's final turn he uses a VCM which is similar in construction to the one used by the student five turns previously, but containing the term 'semantic field'. In this way he can be seen to be signalling that her contribution is valid by ratifying it within the domain of textual analysis as accepted within the academic discourse community.

The data comparisons have shown differences in the realisations and distribution of VCMs across the three data-sets. The analysis has underlined the view that VCMs are highly context-sensitive and reflect the assumed domains of shared knowledge within the three contexts under examination here (informal casual conversation among friends and intimates, radio phone-in and university classes). The conversational VCMs range wide, from those of universal reference to those referring to people and things known only to intimate groups. The Irish radio phone-in data show a great preference for general issues and those of relevance to the national 'community' to which the programme is broadcast, enabling the exchange of views against the background of a socio-political commonage. The academic data are characterised by VCMs which refer to local preoccupations within specific academic disciplines and are concerned with constructing disciplinary understandings and knowledge.

VCMs And Language Teaching

As regards language teaching, a number of pertinent observations may be made:

1   Utilitarian models of language based on transactional premises such as information transfer and information gaps (for example, stronger versions of communicative language teaching) run the risk of stifling the cognitive and linguistic development which is facilitated by a more creative, open-ended approach to language learning. Creativity in all its aspects, not just the more conventional, aesthetic notion of creativity, should be central to language development (Carter and McCarthy 2004). VCMs do 'transfer information', but in quite a different way from the more traditional notion of filling an 'information gap'. Activities in the classroom should be designed to provide space for vagueness and not always seek precision.

2   The lexical realizations of vague categories are, as this chapter has argued, highly patterned and eminently 'learnable'; they are chunks, and fit in well with the lexical approach to language teaching. However, as we have attempted to show, they are also context-sensitive and must always be explored and decoded in context.

3   This chapter has demonstrated that vague categories operate at different levels of assumed shared knowledge: some knowledge can be assumed to be shared by all mature, aware human beings; other knowledge is more locally constrained and culture-bound. Clearly, language teaching has to take the problems posed by restricted references into account, and some teachers may decide initially to

eschew the more circumscribed contexts as a distraction from the vocabulary-learning task, and focus on more universal references. But teachers at higher levels may see restricted contexts as windows on culture and as a site of investigation and potential bridging across cultures, the locus of the third place between the target culture and one's own starting point as a learner (Kramsch 1993).

4   VCMs and their domains of reference are a clear example where corpus insights have an important role to play in informing language teaching materials. As with many high-frequency phenomena in spoken language, intuition, whether that of the native-speaker or the non-native expert user, is likely to be less than adequate to the task of teasing out the commonest expressions, simply because of the real-time, online nature of face-to-face interaction. Language teaching can only benefit positively from the ability of the computer to see large-scale patterns in corpora collected across a range of contexts and users.

Future Research

Several directions for future research emerge from the present study:

1   Spoken corpora need not focus exclusively on the speech of native speakers; comparisons between native-speaker VCM usage and that of non-native expert users will undoubtedly prove equally fruitful, whether in terms of presence or absence of particular lexical types (Prodromou 2005), or in the possible

realizations of the interpersonal functions of vague markers by other strategic means.

2   Learner corpora (especially those coded for errors) are likely to reveal interesting features concerning the successful (or otherwise) acquisition and use of typical VCMs. One frequent phenomenon many second language teachers will be familiar with is lack of concord in expressions such as 'all these kind of things', which are often marked down by teachers and examiners. It remains to be demonstrated, however, that native speakers do not routinely do the same thing, and casual observation of the CANCODE corpus suggest that native-speaker examples such as 'the property was being kept clean and tidy and all <u>those kind</u> of things' are by no means rare.

3   Research into processing, involving protocols and similar methods, may reveal much about how learners or non-native users (or any group of outsiders) process and decode the referents of VCMs. Here corpus observations and more psycholinguistically oriented research can fruitfully contribute to each other (for example see Spöttl and McCarthy 2004 on formulaic sequences).

4   Corpus-based cross-linguistic comparisons of VCMs are needed, especially for lesser-researched languages, both in terms of syntax and semantics and pragmatics. Models based on high-attention languages such as English tend to dominate; research examining other languages in their own right can serve to ratify or challenge English-dominated models. The same applies to varieties within languages such as English, where certain varieties tend to have dominated, though in the case of British and Irish English, the balance is

swinging into greater equilibrium. This chapter hopes to make a contribution to that effort.

References

W.L. Barsalou, 'Ad Hoc Categories', *Memory and Cognition,* 11 (1983) 211-77.

W.L. Barsalou, 'The Instability of Graded Structure: Implications for the Nature of
 Concepts', In U. Neisser, (ed) *Concepts and Conceptual Development*,
 (Cambridge: Cambridge University Press, 1987), p.101-40.

R.A. Carter and M.J. McCarthy, 'Talking, Creating: Interactional Language, Creativity
 and Context', *Applied Linguistics*, 25/1 (2004) 62-88.

R.A. Carter and M.J. McCarthy, *The Cambridge Grammar of English*, (Cambridge:
 Cambridge University Press, 2006).

J. Channell, *Vague Language,* (Oxford: Oxford University Press, 1994).

P.D. Crawford, 'Educating for Moral Ability: Reflections on Moral Development Based
 on Vygotsky's Theory of Concept Formation', *Journal of Moral Education,* 30/2
 (2001) 113-129.

J. Cutting, *Analysing the Language of Discourse Communities,* (Oxford: Elsevier
 Science, 2000).

E. Dines, 'Variation in Discourse - and Stuff like that', *Language in Society,* 1 (1980) 13-
 31.

S. DuBois, 'Extension Particles, etc.', *Language Variation and Change,* 4 (1993) 179-203.

E. Goffman, *Forms of Talk,* (Oxford: Basil Blackwell, 1981).

G. Jefferson, 'List Construction as a Task and Resource', In G. Psathas (ed.) *Interaction Competence,* (Lanham, MD: University Press of America, 1990), p.63-92.

A.H. Jucker, S.W. Smith and T. Lüdge, 'Interactive Aspects Of Vagueness in Conversation', *Journal of Pragmatics*, 35 (2003) 1737-69.

C. Kramsch, *Context and Culture in Language Teaching*, (Oxford: Oxford University Press, 1993).

R.K.S. Macaulay, *Locating Dialect in Discourse: the Language of Honest Men and Bonnie Lasses in Ayr*, (New York: Oxford University Press, 1991).

M.J. McCarthy, A. O'Keeffe and S. Walsh, ' "Post-colonialism, Multi-Culturalism, Structuralism, Feminism, Post-Modernism and so on and so forth" – Vague Language in Academic Discourse, a Comparative Analysis of Form, Function and CONTEXT', Presented at AAACL/ICAME conference, University of Michigan, Ann Arbor, (2005).

A. O'Keeffe, 'Exploring Indices of National Identity in a Corpus of Radio Phone-In Data from Irish Radio', In A. Sánchez-Macarro, (ed.) *Windows on the World: Media Discourse in English*, (Valencia: University of Valencia Press, 2002), p.91-113.

A. O'Keeffe, ' "Like the Wise Virgins and all that Jazz" – using a Corpus To Examine Vague Language and Shared Knowledge', In U. Connor and T. A. Upton, (eds.) *Applied Corpus Linguistics: a Multidimensional Perspective*, (Amsterdam: Rodopi, 2003), p. 1-20.

A. O'Keeffe, *Investigating Media Discourse,* (London: Routledge, 2006).

M. Overstreet and G. Yule, 'On Being Explicit and Stuff in Contemporary American English', *Journal of English Linguistics*, 25/3 (1997a) 250-58.

M. Overstreet and G.Yule, 'Locally Contingent Categorization in DISCOURSE', *Discourse Processes,* 23 (1997b) 83-97.

L. Prodromou, 'You See, it's Sort of Tricky for the L2 user': the Puzzle of Idiomaticity in English as a Lingua Franca. Unpublished PhD dissertation. University of Nottingham, UK (2005).

M. Scott, *Wordsmith Tools* (software), (Oxford: Oxford University Press, 1999).

C. Spöttl and M.J. McCarthy, 'Comparing the Knowledge of Formulaic Sequences Across L1, L2, L3 and L4', In N. Schmitt (ed.) *Formulaic Sequences*, (Amsterdam: John Benjamins, 2004) p.191-225.

J. Swales, *Genre Analysis: English in Academic and Research Settings*, (New York: Cambridge University Press, 1990).

L.S. Vygotsky, *Thought and Language*, (Cambridge, MA: The M.I.T. Press, 1962).

L.S. Vygotsky, *Mind in Society*, (Cambridge, MA: Harvard University Press, (1978).

G. Ward and B. Birner, 'The Semantics And Pragmatics of "and Everything"', *Journal of Pragmatics,* 19 (1992), 205-214.
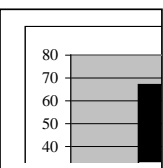
Notes

---

[1] CANCODE stands for 'Cambridge and Nottingham Corpus of Discourse in English'. The corpus was developed at the University of Nottingham, UK, and funded by Cambridge University Press, UK, with whom sole copyright resides. CANCODE forms part of the larger Cambridge International Corpus. The corpus conversations were recorded in a wide variety of mostly informal settings across the islands of Britain and Ireland, then transcribed and stored in computer-readable form. Details of the corpus and its design may be found in McCarthy (1998).

[2] Source: JNLR/MRBI radio figures released February 2003, quoted in Oliver (2003).

[3] *For the X that's in it* is a dialectal form found in Irish English but not familiar to most British English speakers. It is a direct translation from Gaelic, and marks a vague category relating to special occasions and the activities and behaviors which are associated with them. For example, on someone's birthday *we're having a get together for the day that's in it,* or on the millennium year, *there were lots of celebrations for the year that was in it*.

[4] Irish secondary schools are non-fee paying, but there are still some which have a boarding facility for which fees are paid and students stay at the school during the week, or for longer periods.

[5] An Irish car hire company.