

‘Like the wise virgins and all that jazz’ – using a corpus to examine vague categorisation and shared knowledge

Anne O’Keeffe

Mary Immaculate College, University of Limerick, Ireland

Abstract

This paper will use a corpus to explore vague categorisation (e.g. prostitutes, sailors and the like) in a specific context where the participants are strangers, but where they share the same socio-cultural reference points and so can assume a critical level of shared socio-cultural knowledge when they use vague language. Unlike most work on vague language, this study looks at vague items which are not necessarily pre-textual or prototypical, but which emerge from shared knowledge. The data comprises 55,000 words of calls to an Irish radio phone-in show. Vague category markers are isolated and described in terms of form and domain of reference. It is argued that the shared knowledge required in order to construct vague categories has a common core of socio-culturally ratified ‘understandings’ and that the range of domains of reference of these categories is relative to the depth of shared knowledge of the participants and relative to their social relationship.

1. Introduction

Much theoretical debate surrounds the epistemic (i.e. knowledge) status of *vagueness*. According to the epistemic theory of vagueness, there is no absolute state of ‘borderline’. If someone is borderline bald, for example, this theory holds that s/he is either definitely bald or not bald, but we (as the speakers) cannot at that point determine (see for example, Williamson, 1994). However, Jackson (2002) argues that the role of language in communicating our thoughts about how things are makes a strong case against this absolutist theory. Other recent philosophical arguments look at vague language in context (see for example Pinkal 1985; Manor 1995; Kyburg and Morreau, 2000). Kyburg and Morreau (2000), for example, take the stance that ‘contextuality’ and ‘accommodation’ are characteristic of vague language between speakers and hearers in context:

...just as a handyman can fit an adjustable wrench to a nut, we think, a speaker can adjust the extension of a vague expression to suit his needs, relying on the hearer to recognize his intentions and to accommodate him.

Kyburg and Morreau, (2000: 577)

The linguistic study of vague language has been greatly influenced by Grice's (1975) Co-operative Principle (CP) and its associated conversational maxims. Most notable is the work of Cruse (1975 and 1977) who points to the relativity of vagueness: 'a speaker wishing to refer to something in his surroundings is frequently, if not usually, faced with a range of lexical items of different levels of specificity, all of them equally appropriate from the point of view of their inherent sense' (Cruse 1977: 153). Cruse (1977) explores the notion of unmarked or neutral levels of specificity in various contexts which are not necessarily covered by Grice's maxims. He presents a system of markedness in terms of level of specificity. Of salience to the present study are some incidental comments made by Cruse in this 1977 paper. Firstly, he makes the point that under-specification de-emphasises the feature that is omitted, while over-specification emphasises or intensifies the added feature (ibid: 163) as an example of under-specification, he mentions expression of compassion or pity. Apart from under-specifying simply for reasons of unwillingness to give information, Cruse also points out that a speaker may underspecify because s/he is an expert in a particular field, or has at least an everyday familiarity with some class of things: 'the speaker is in effect suggesting that the referent has such a high degree of givenness in his universe, that he cannot make what is an unmarked reference without underspecifying' (ibid).

It is this notion of 'givenness' which Cruse associates with vague language use that is of interest in this paper. We will examine the use vague categorisation in a very self-contained context where speakers within the same society draw on their shared knowledge in the frequent use of vague categorisation. The data will be taken from a small corpus of radio phone-in data from an Irish radio show called *Liveline* (see 3 below). When the *Liveline* presenter and callers underspecify, they are drawing on assumptions and expectations about the 'givenness' of the shared social and cultural knowledge and information of their co-participants. Take the following example where a caller detailing her experience of Maori body tattoo draws on the givenness of our knowledge of the human physique:

1)

Caller: And their tattoos were absolutely weepingly beautiful. They were extraordinary. And those men were tattooed all that I could see okay so starting with **the forehead face ears neck hands et cetera et cetera.**

Presenter: Yeah.

Here the caller can take a linguistic shortcut using the vagueness marker *et cetera*. This allows her to say 'the forehead face ears neck hands et cetera et cetera' instead of forehead, face, ear, neck, hands plus a tiresome list of all bodily parts that were tattooed. Our shared knowledge of the human body combined with the speaker's knowledge of the givenness of this information facilitates such under-

specification. At a more culture-specific level, we find the following type of example in the data where the ‘given’ or implicit information is not as universal as in the previous example:

2)

A caller is reminiscing about his schooldays in an Irish boarding school fifty years ago.

Caller: ...you were supposed to be on a rugby pitch **or something like that** you know ...

Presenter: Right.

Here, in order to complete the referential set *a rugby pitch or something like that*, the listener needs to have shared information from an Irish social context of the type of games that are usually played in an Irish Catholic boarding school fifty years ago. An ‘outsider’ (i.e. someone from outside of Irish society) hearing this utterance can engage with it to a certain degree, either by under-specified broad human knowledge, or by (possibly incorrectly specified) cultural analogy with his/her own culture.

The aim of this paper is to examine vague categorisation in context using a self-contained corpus of data as a measure of the range of shared or given information of the participants. In other words, by isolating and analysing all of the vague categories that are constructed by the speakers in the data, it is hoped to find indexical information about these participants. In so doing, we may gain an insight into the nature of the shared knowledge that binds this group. It will also be argued that a corpus provides a very useful tool for the study of vague language in use.

2. Previous research

Vague language is defined in a number of ways. Franken (1997) distinguishes between ‘vagueness’ and ‘approximation’ while Channell (1994) restricts the definition of vagueness to ‘purposefully and unabashedly vague’ uses of languages. She divides vagueness into three categories: 1) vague additives (which include vague approximators such as *about* and tags referring to vague categories such as *and things like that*) 2) vagueness by choice of words (e.g. *yoke*; *thingy*) and vague quantifiers (e.g. *piles of*) and 3) vagueness by implicature (e.g. the sentence *Sam is six feet tall* has the potential to be vague as he may be six feet and a quarter of an inch tall; see Channell, 1994: 18). On the other hand, Zhang (1998) makes a case for four separate categories: ‘fuzziness’, ‘generality’, ‘vagueness’ and ‘ambiguity’. Unlike Zhang, Chafe (1982) puts vagueness and hedging together into the 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 stance taken here: that to be vague is to draw on what is given and shared within the participation framework of the Irish radio audience.

Similar to Channell (1994), Powell (1985) focuses on the notion of purposeful vagueness. She deals with vague quantifying expressions and argues that 'a maximally efficient exchange of information may be vaguely encoded, and purposively so, if the principal function of the exchange is essentially non-descriptive' (ibid: 32). She also shows that vague quantifying expressions may encode a speaker's judgement and that this dimension of use is principally evaluative in function. The following example from the radio data clearly fits this model.

3)

Presenter: Why did you decide on boarding school?

Caller: Well we live in the country and the nearest school to us was going to be fifteen miles away where we= our boys would be big into **sports and all that**.

The presenter asks a straightforward referential question as to the caller's motives for sending her children to boarding school. The non-descriptive answer gives two motives 1) the distance from the school and 2) her children's love of school team sports and school activities which might not have been sated had they lived at home given the distance of the family home from the school and the need to stay on after school hours for games, training and school outings. Here we see that 'our boys would be big into sports and all that' serves as a shortcut to motive number 2 above. What is of note for this investigation of vague language in use is the level of assumed knowledge anticipated on the part of the caller in using this vague linguistic shortcut. This reference is ratified by the presenter who finds the caller's explanation adequate and unambiguous (either for her or for the listeners on whose part she arbitrates). It is also interesting to note that when the caller says 'we live in the country', this is implicitly understood within an Irish context. It would be taken to mean that *we* are farmers living outside a town or village away from a school bus route. In another social context this would often carry different implications.

2.1 Categories and categorisations

In this paper, analysis will be limited to vague category markers as such as those illustrated in examples 1, 2 and 3 above (as opposed to looking at individual vagueness items such as nouns, quantifiers and so on). *Vague category markers* go by various terms across different studies for example, 'general extenders' (Overstreet and Yule 1997), 'generalized list completers' (Jefferson 1990), 'tags' (Ward and Birner 1992), 'terminal tags' (Dines 1980; Macaulay 1991), 'extension particles' (DuBois 1993) and 'vague category identifiers' (Channell 1994).

Most research into the nature of categories has been concerned with lexicalised categories, that is those that are encoded as a single lexical item (for example, *bird*, *furniture*) see for example Mervis and Rosch (1981); Rosch (1978) and Rosch, Mervis, Gray, Johnson and Boynes-Braem (1976). Many of these studies look at categories in terms of prototypes (exhibiting the highest concentration of characteristic properties) compared with peripheral category members (containing fewest characteristic properties). Of more relevance to this study, Barsalou (1983 and 1987) looks at the question of whether categories are stable or subject to change. In particular he 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. This challenges the notion that categories are stable easily recognisable and arrived at 'pre-textually' (after Overstreet and Yule 1997). Overstreet and Yule (1997) 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.

Overstreet and Yule (1997: 85-6)

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 (1997: 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. They refer to the set of forms that generate non-lexicalised categories as 'general extenders' which they see as integral to the process of establishing categories that are locally contingent in discourse. In this analysis these forms will be termed 'vague category markers'.

The vague category markers in the corpus will be seen as recognisable chunks of language that function in an expedient way as linguistic triggers employed by speakers and decoded by participants who draw on their store of shared knowledge. It is argued here that the meanings of vagueness categories are socio-culturally grounded and are co-constructed within a social group that has a shared socio-historic reality. However, it is wise to issue the caveat that without access to the speakers for personal reflection, we cannot know for certain whether they choose to take linguistic shortcuts: a) to be 'deliberately and unresolvably vague' (Powell 1985: 31) or b) to be expeditious and adhere to conversational norms of quantity.

3. Data

Data for this study is drawn from an Irish radio phone-in show called *Liveline* broadcast every weekday on Radio Telefís Éireann (RTÉ) between 1.30pm and 2.45pm. The transcribed corpus comprises approximately 55,000 words. The programme has been running for the almost 18 years and according to recent research has an audience of 365,000¹, almost 10% of the Irish population. Its longevity and prominence on Irish airwaves makes it rich for analysis on many levels. The data was taken from a sample of programmes in 1998, and comprises 44 phone calls (from a total of five programmes) spread throughout that year. Programme selection dates were spread throughout the year at intervals that would avoid daily or seasonal skewing (i.e. spread around different days of the week and months of the year at more or less equal intervals). Once dates were chosen, the relevant programmes were recorded from the RTÉ radio archive and the researcher had no prior knowledge of what topics would be covered on these programmes. In the data, topics for discussion meander from call to call and include the following miscellany: female facial hair problems; tattoos; the peace process in Northern Ireland; how ears were pierced in the old days; constitutional referenda, experiences of working aboard; cursory tales about sunbathing without sun block; reminiscing about boarding schools; warnings about the decline of fidelity and moral decay in general; things that can go wrong when working in Saudi Arabia and the growing trend of litigation in Irish society among others. Unlike many talk radio shows, the presenter in *Liveline* does not normally provide counselling and she generally avoids engaging in strong debate. Her role appears to be more that of conduit between the caller and the audience (see also O’Keeffe 2002, McCarthy and O’Keeffe 2003).

4. Analysis

The analysis focuses on *any* forms that make vague reference to sets or categories. Research tells us that vague category markers are found in clause-final positions and mostly comprise a conjunction and a noun phrase however because a bottom up approach to identifying all vague categories in the data was used, there was no pre-selection criteria based on form. This poses a challenge for corpus analysis in that this data must first be checked manually. However, since we are dealing with a small corpus, this is not an impossible task. While corpus tools assist in checking the accuracy of the manual searches, there still remain questions of validity and reliability. In order to enhance these aspects of the study, two raters were used. One of the raters was from the Republic of Ireland and one from Great Britain. Retrospectively, it proved very important to have one rater who was familiar with the cultural references in the data. However, the non-

¹ Source: JNLR/MRBI radio figures released February 2003, quoted in Oliver (2003).

Irish rater proved to be crucial to validity and reliability of the study as he was in a position to see exactly when a vague category was exclusively within an Irish reference domain (see section 5.2). Surprisingly, it was not always as easy for the Irish rater to see the range limit in her own cultural reference domain. Such a study could not be conducted without this inter-rater reliability check.

In this analysis therefore, the 55,000 word corpus was searched exhaustively by rater 1 to identify manually any forms that were used in vague categorisation (as a follow up, *Wordsmith Tools* used to generate accurate quantitative results). These were cross checked by rater 2. The categories or 'sets' which were found in the data were then logged for subsequent analysis. Any forms which were co-textual, i.e. referring back to a previously identified set or category made explicit in a previous turn, were not included - see extract 4 below as an example where this is the case. The form *anything like that* appears to be a vague category marker, but on closer examination, it refers back to a catalogue of headaches which the caller details earlier in the turn:

4)

A caller talking her experiences of the side effects of taking a contraceptive pill.

Caller: Am well I'd nasty headaches very nasty headaches am I was on it for a month. I went on to it for the second month and a couple of days into it I was out one night I wasn't feeling myself and I went home and the following day I'd ah very very bad headaches and I knew there was something wrong myself because I'd never experienced **anything like that** and I don't suffer from migraine so am I went to bed got up on Monday went to work felt dreadful in myself as well as having the headaches+

5. Results

In all, 138 vague category markers were identified. Each form was classified according to its reference set as the following example shows for the set: *a lot of undesirables criminals and people like that...* (while this might appear to be a global reference, it was used in a specifically Irish context by the caller):

Table 1 – Sample breakdown of initial analysis of categories

Form	<i>a lot of undesirables criminals and people like that...</i>
Reference set <i>(i.e. the set or category that is alluded</i>	<i>Criminals and social</i>

to)	<i>undesirables</i>
Broad category (i.e. the broad category of reference, e.g. Irish historical, global etc.)	<i>Irish social</i>
No. of occurrences	2

5.1 Forms of vagueness

The following distribution of forms was identified in the *Liveline* data.

Table 2 - Distribution of vagueness forms used to mark categories in *Liveline* data²

Form	Example	Raw result	Result per million words
thing(s)	we're going to get a clatter of phone calls talking about there was one nurse I can't remem= <chuckles> was it nurse Caddin wasn't she involved in the most extraordinary things in Dublin?	46	836
X like that	a lot of undesirables criminals and people like that...	21	381
...[that/that] [kind/sort/type] of X	unhappy homes all that kind of thing	17	309
and so on	conviction about social justice and so on	12	218
Or any/something	And it worked very well in fact the day boys were very useful because you could often get them to get fags for you up town or [Yeah] buy a bar of	9	164

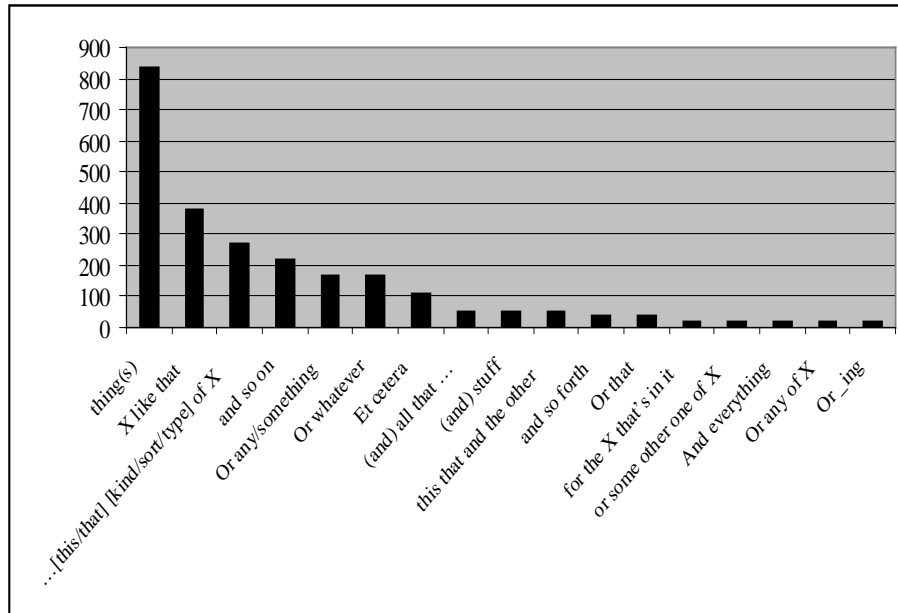
² Round brackets indicate lexical items that may co-occur and items within square brackets are alternative but mutually exclusive (e.g. *that [kind/sort/type] of X* implies *that kind* or *sort* or *type of X*).

	chocolate or something and smuggle it in you know.		
Or whatever	the expense of insurance and ah people for instance organising voluntary sporting activities now find that you know if you have a gymkhana or whatever ...	9	164
Et cetera	the development of piers, roads et cetera and et cetera	6	109
(and) all that ...	maybe they are like the w= the wise virgins and all that jazz	3	54
(and) stuff	...out at discos and stuff	3	54
this that and the other	with this that and the other thing	3	54
and so forth	talking about married men an= ma= and so forth	2	36
or that	I didn't know anything about lights or that and they told me that the lights was affecting his eyes	2	36
for the X that's in it	a bit of respect for the day that's in it.	1	18
or some other one of X	the ozone layer or some other one of these quare things up there in the sky	1	18
and everything	the whole attitude in the school is like rugby at the moment it's the rugby season and the Cup and everything ...	1	18
or any of X	I'd like Bertie or any of them get on and address what we're voting on on Friday	1	18
Or _ing	. I just saw a lot of kids now by kids I mean up to maybe age of twenty-four or so [Mm] enjoying themselves or doing whatever they do to that particular form of sound they use as music.	1	18

TOTALS	138	2.505
---------------	------------	--------------

These results are presented in fig. 1 below:

Figure 1 – Forms of vague category markers found in the *Liveline* data (occurrences per million)



By including any form in the data that is used to construct a vague category, we find a somewhat unorthodox collection. As mentioned above, Channell (1994) noted that most vague category markers were clause-final (conjunction) + noun phrase pattern; however, the results here deviate from this, for example adverbial phrases *and so on/and so forth*, with no noun phrases. In particular, we also note the inclusion of *thing(s)* which is not necessarily clause final. For example:

Figure 2 – Sample concordance lines for *thing(s)*

2	gs just won't laughs am some things have to be faced
3	with rose tinted spectacles and saying things were great you don't
4	st I mean I don't mean to be dismissive things like social justice I
5	want certain things to go away but some things just won't
6	le but it's um an Islamic country. It's things are so so so
7	eah. +which are doing different things and I can be left out
8	yes ah of doing unspeakable things to one another ah to
9	before they were in a rural area I mean things have just changed.

10	roller coaster you know I mean I think	things are moving very fast
12	Oldest swinger in town kind of thing?	Absolutely.
13	emo tour of Italy and all these kind of things	about four years ago+
15	Which is dependent upon all kinds of things	happening in the
16	he whole idea is that a fresher look at things	ah by looking at the
17	I did fifty five years ago and a lot of things	have passed under the
18	RU and they get criticised for a lot of things	I don't think it was
19	get to understand that you owe a lot of things	to other people in
20	in the most extraordinary things	in Dublin?
21	so yo= you know there's a lot of funny things	happening in boarding
22	rse it does but she hears a lot of good things	about it too. It's
23	it's associated with all sorts of seedy things	like venereal
25	sound mild in comparison to some of the things	that other people

Fronek (1982), writing on the word *thing(s)* notes that 'the poverty of its semantic content makes it a very good candidate for the various degrees of semantic re-categorisation and for use as a function word' (Fronek 1982: 636). Many linguists (see Bloomfield 1933; Hockett 1958; Gleeson 1956 and Lybbert 1972) have focused on the capacity of the word to acquire grammatical function because of its 'semantic emptiness' (Fronek 1982: 636). Fronek illustrates that there is no sharp distinction between the lexical and grammatical classes. *Thing* is an extremely flexible function word capable of shedding most of its semantic content and thus becoming suitable as a pro-form while also capable of behaving like any other noun (ibid: 652-3). However, Fronek notes that 'especially the plural indefinite *things* can have such a vague indeterminate referent as to be almost indefinable. Its notional content is so minimal that from the semantic point of view it is virtually redundant' (ibid: 645). This assertion is disputed here, at least in the context of the current study, where it is held that what might seem to the analyst to be indeterminately vague is communicatively and pragmatically adequate to the collective users in context. Take the following example:

5)

The caller is talking about a boarding school he attended many years ago and he has just mentioned that at one point it became a mixed gender school.

1. Caller: ... that was sort of a <unintelligible word> an indicator of what **things** were to come in the future you know.
2. Presenter: Yeah yeah. Well I mean there now in a very built-up area whereas before they were in a rural area I mean **things** have just changed.
3. Caller: Th= that's right. Yeah that's right.

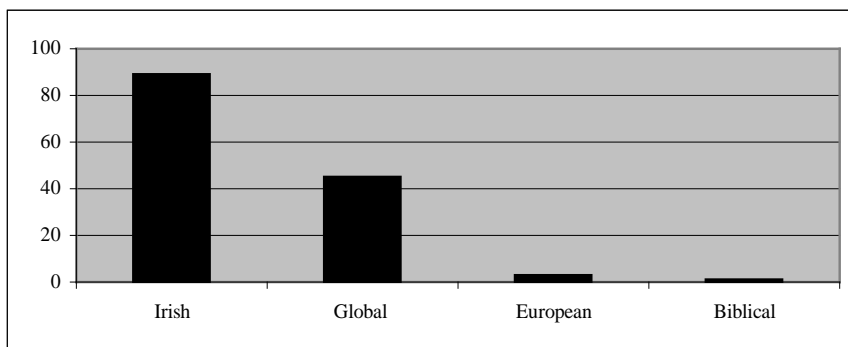
The first use of *thing* (by the caller in line 1) refers to *things* that were to change in relation to the school and the presenter ratifies her understanding of this with *yeah yeah* (line 2). The presenter then uses *things* (line 2) to refer to broader

changes in the area around the school (Newbridge College, situated in County Kildare outside Dublin), which fifty years ago was rural and which now is a very built up satellite town of Dublin. To an non-Irish observer, it is fair to say that these two uses of *thing* are almost opaque, but to someone who has access to the social information of the participants, this is a normal inexplicit reference to given, shared information.

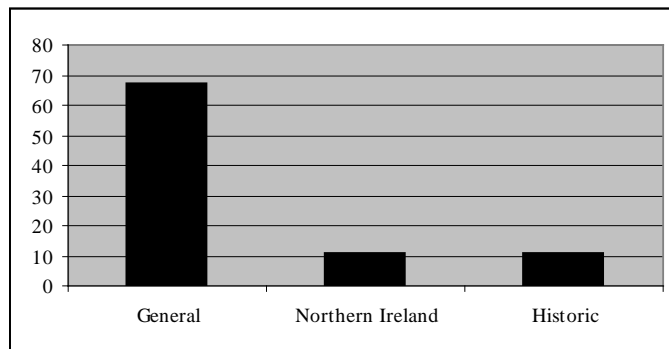
5.2 Reference domains

The 138 vague language forms were distributed across three broad domains: national references (i.e. Irish), global, European, and a fourth, minor category, biblical. Rater 1 sorted the items into these broad categories and rater 2 cross checked these. As discussed above, rater 2 was from outside of Ireland and was better placed to identify solely Irish references.

Figure 3 - Broad reference domains of categories



As we can see, these sets fall mostly into two reference domains: *Irish* and *Global*. A further breakdown of the *Irish* category is profiled below:

Figure 4 - Breakdown of the reference domains at a national level (i.e. *Irish*)

The *General* set is all references that are not related to Northern Ireland issues or historic collective Irish knowledge. They are contemporary social reference points spanning a multitude of social issues and information. This could be seen as the most core or most common information held within the group of participants. Examples from this category include: typical accidents that happen to people in Ireland; small midland towns in Ireland; typical contemporary issues that are discussed in Ireland; social activities typical of an Irish teenager.

5.3 Categorisation as generic indexical information

The categories co-constructed within the participation framework of *Liveline* give clues as to the profile of the audience. Clearly, it is an Irish-centred one, with the main core of reference points centring on general Irish social knowledge. When this 'general' data is scrutinised more closely with the help of concordance line analysis, we find that the locally contingent categorisation can be divided into four categories 1) social practices and attitudes; 2) social responsibilities and realities; 3) work, financial and consumerist practices and 4) social types. Raters 1 and 2 devised these four categories and then independently sorted the items. Their results were then compared and any anomalies were scrutinised and resolved. Table 3 provides an example for each of these:

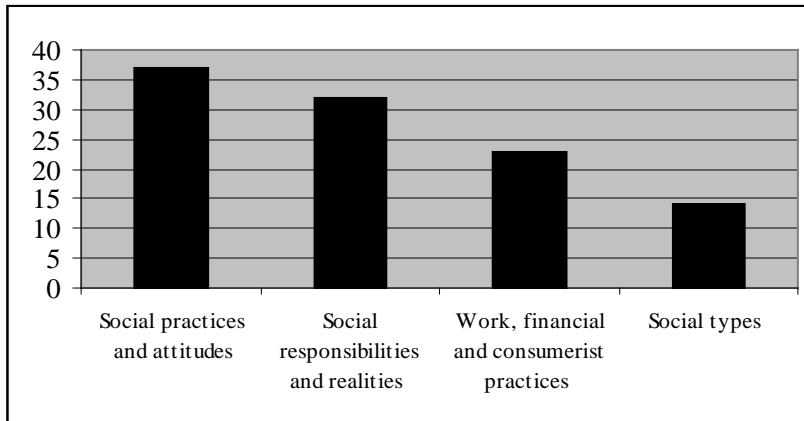
Table 3 – Categories within in *General* Irish reference domain

Category	Example
Social practices and attitudes	The process of 'word of mouth' in Ireland
Social responsibilities and realities	Negative social realities that come with the Celtic tiger economy

Work, financial and consumerist practices	Car rental companies in Ireland
Social types	Irish criminals and social undesirables

Figure 5 shows how these are distributed in the corpus of data:

Figure 5 – Percentage breakdown of general Irish references



Based on these core reference points (i.e. most general or ‘common denominator’), it is fair to assert that they index or place the participants of the radio phone-in show *Liveline* as a socially-aware, middle class group. Most telling in this respect are the frequent categorisations in relation to ‘others’ in Ireland who are socially disadvantaged (for an in-depth treatment of this see O’Keeffe 2002).

5.4 Stereotypes and prototypes

As discussed earlier, much research has looked at semantic prototypes in the construction and stabilising of categories. However, it is of note that when we look at their construction within the stable participation framework of *Liveline*, we find that many of the social references are dependent on (and symptomatic of) stereotyping. In example 6 below we find a typical example:

6)

Talking about why people send their children to boarding schools.

Presenter: And you I mean ab= some people were there saying oh well sending them away **unhappy homes all that kind of thing**

For the majority of people listening to this caller at the time, we can only assume that they deconstruct the meaning of the category *unhappy homes all that kind of thing* based on a stereotype as opposed to direct first-hand experience and this is the case for many of the examples found in the data.

5.5 Categorisation and semantic prosody

As noted above, Cruse (1977) tells us that under-specification de-emphasises the feature that is omitted. On examination of the semantic prosody of the categories constructed in the data, we find many that are negative. It could be posited therefore that one of the motivations for using vague categories could be to avoid over-specification in negative domains. In other words, many of these uses could serve as euphemisms. Within the context of the work of Louw (1993) and Sinclair (1996) on semantic prosody, we could say that vague category markers cluster with lexis which has negative prosody. The degree to which the under-specification of negative categories is culturally marked cannot be measured or proven here but it is put forward that it may be so. Fig. 6 below provides a sample of some of the negatively marked lexical items which collocate with vague category markers in the data:

Figure 6 – Samples of negative semantic prosody associated with categorisation

4	ell sending them away unhappy homes all that kind of thing.	Mm I know
6	mean there was an awful lot of pain and that kind of thing	.
7	ools are from unhappy families there is that kind of element I suppose but mo	
8	ke venereal diseases or prostitution or that kind of thing?	Well I I
10	der. And had he been subject to that kind of physical torture ?	
11	n't I. And you know Marian if you're in that kind of am hostile environment	
18	ld is going to be and so on I mean that that sort of issue I think we need to	
19	create divisions and conflicts and all that sort of thing.	Yeah.
20	react quite strongly to stress and all that sort of stuff so I have I'm now	

2	g out of this ah situations of hardship and so on I think we wouldn't say tha	
4	xecuted and the other was to get lashes and so on and so forth. Yo= yo= did y	
10	ing through ah this system and the pain and so on and so on. But having said	
15	h ah this system and the pain and so on and so on. But having said that let	
	me	

112	it's associated with all sorts of seedy things like venereal diseases or	
120	Won't that be the most subversive thing that has been done to both sets	
123	e the point of road rage ah this is the thing I'm concerned with where local	
132	or you know the the the danger is these things get worse and worse and become	
133	t okay so let people be beware of these things. Okay?	Yeah. But the
134	h am tough-minded view would say "these things happen. It's too bad".	
139	roller coaster you know I mean I think things are moving very fast but all	

151	yes	ah of doing unspeakable things to one another ah to come
-----	-----	---

6. Categorisation and hedging

It could be argued that the construction of vague categories serves as a form of hedging as the following example illustrates, where the presenter is asking a difficult question, that is whether the caller was in receipt of a government ‘hand out’ (note: the *Gaeltacht* refers to areas of Ireland where Irish is the first language).

7)

Presenter: Didn't get a Gaeltacht grant **or anything like that?**
 Caller: No I didn't get anything not a grant aid whatsoever.

The vague category marker clearly functions to downtone the accusation implicit in the question. We posit that when speakers want to hedge the force of negative utterances they can choose to construct a vague category as a discourse strategy. This supports Cruse's point cited earlier that under-specification de-emphasises the feature that is omitted (cf. Cruse 1977).

7. Categorisation as a generic activity

Warren (1993) tells us that inexplicitness (of which the construction of vague categories is one exponent) depends on overlapping factors: 1) the physical setting and 2) shared knowledge. Transposed onto a national radio audience level, this assertion fits the findings in this study where the majority of the vague categories constructed have their reference domain in physical or social space in Ireland, and all are bound by an almost uncontested ratification by participants in the construction of their meanings. In other words, these vague categories are perfectly transparent for their users (though this may not be the case for the analyst). This has interesting implications for the study of spoken genres. It points to the speaker-addressee interdependence in the co-construction of meaning and it points to the bi-directionality of spoken discourse. Take the following example:

8)

1. Caller: I have Emm she's fourteen and her brother slags
 her now he's sixteen he would be going ‘look
 you have you have hair unde= you have a

moustache” **and all this** so I do have to give out to him.

2. Presenter: Yeah.

A non-lexicalised category of things that an Irish teenage boy might say to tease his sister who has a facial hair problem (and even how it might be said) is vividly invoked by the caller with minimal lexical effort: *he would be going “look you have you have hair unde= you have a moustache” and all this*. This is perfectly understood by the addressee (and we assume by the audience as hearers), but crucially it is facilitated by the triangulation between all three: speaker; addressee and hearer(s) because they know the range of common knowledge that the other knows.

Over time the participants develop a sense of the internal range of shared knowledge which can be drawn on. In other words, the range of shared knowledge accrues within the participation framework. This store of shared information allows speakers to draw on generic resources with minimal lexical effort. Consider the following example:

9)

A caller who owns a hostel in the West of Ireland is telling a cautionary tale about a man who pretended to be a member of staff and who stole some guests’ luggage. Note: *Gardaí* refers to the Irish police force.

Caller: ...in the hostel Marian there’s one very clear practice with hostellers and that is honesty and trust in one another.

Presenter: Yeah.

Caller: They would not take a simple tea bag unless they ask for it. Not one.

Presenter: Right. Okay.

Caller: And once this trust is broken down hostels will cease to exist.

Presenter: Okay well I suppose it is fair to say that am the Gardaí could pursue it but I guess ... I don’t know ho= how the decision is made. ... in the order of things the people weren’t that offended et cetera et cetera et cetera and there are drug barons **et cetera** out there **you know yourself** how the argument goes ...

Here we find generic activity where the presenter can invoke a whole line argument through the delexicalised category: ‘there are drug barons **et cetera** out

there **you know yourself** how the argument goes'. Here, she is drawing on the shared knowledge of the caller and the audience that in the Irish media there has been much debate as to how the *Gardaí* should allocate resources, for example, whether they should prioritise serious crime issues such as drugs and criminal gangs or whether they should invest more in basic safety for the average citizen by following up on smaller crime issues such as this one. This is again an example of the dynamic and collaborative nature of spoken discourse and how dependent it is on its physical and social contexts and the shared knowledge of its participants.

8. Conclusion

Analysis of the vague categories in this paper supports Barsalou's (1983) assertions about the dynamic nature of ad hoc categorisation as well as Overstreet and Yule (1997), who stress the spontaneity of categorisation and the context-dependent nature of the categories themselves. By looking at the reference domains of the vague categories which are used by the participants, we find that they index a substantial pool of shared knowledge at an Irish societal level. At a broader level, this supports Bakhtin's criticism that many models of linguistic analysis have failed to understand the nature of utterances because they adopt a passive model of meaning and understanding. They perceive language as a speech flow from the speaker to a passive recipient instead of recognising the active role of the *other* in the process of speech communication (translations of Bakhtin's work in Morris 1994: 80).

The range of shared 'core' knowledge which was identified suggests that the participants draw mostly on shared societal information in the context of a national radio phone-in show and that this reflects their social relationship. The participants are strangers to each other but they do share much societal common ground which allows them to construct vague categories which are mutually understood. It is also argued that this process in its turn helps to create and sustain the pseudo-intimacy that is required for this type of radio interaction, where it is important for callers to feel part of a group that has a shared socio-cultural background.

At a methodological level, this paper shows how corpus linguistic methods can be used to support the initial detailed examination of forms in a small corpus. By taking a manageable amount of data and scrutinising it for all forms that invoke vague categories, we have been able to identify a wider range of forms than if we had approached it with a preset list of form types. The benefit of this bottom-up approach is that we have been able to identify all of the vague categories in the data as well as all of the forms used to construct them in this specific context. Now that the forms of vague categorisation have been identified and analysed in

this corpus of data, it will be possible to compare them to other corpora. For example preliminary research by McCarthy and O’Keeffe (2002) looked at some of these forms in relation to a sub-corpus of the Cambridge and Nottingham Corpus of Discourse in English (CANCODE) which comprised 55,000 words of casual conversation data from close friends. Preliminary findings show that certain forms are not as frequent in casual conversation between friends and that the reference sets in the CANCODE data indexed much narrower ranges of shared knowledge (for example within sub-cultural groupings and workplace shared knowledge). In other words, a correlation between range of shared knowledge and speaker relationship was evident.

Overall this study has shown us that by looking at a corpus of language in use within a particular social context one can access indexical information through patterned use of the language. In this case, we have been able to show that vague language categorisation is a by-product of a close relationship at a societal level of interaction.

References

- Barsalou, L. (1987), ‘The instability of graded structure: implications for the nature of concepts’, in: U. Neisser (ed.) *Concepts and conceptual development*. Cambridge: Cambridge University Press, 101-140.
- Barsalou, L. (1983), ‘Ad hoc categories’, *Memory and Cognition*, 11: 211-277.
- Bloomfield, L. (1933), *Language*. New York: Holt, Rinehart and Winston.
- Chafe, W. (1982), ‘Integration and involvement in speaking, writing, and oral literature’, in: D. Tannen (ed.) *Spoken and written language: exploring orality and literacy*. Norwood, N.J.: Ablex Publishing Corporation, 35-53.
- Channell, J. (1994), *Vague language*. Oxford: Oxford University Press.
- Cruse, D.A., (1975), ‘Hyponymy and lexical hierarchies’, *Archivum and linguisticum* VI: 26-31
- Cruse, D.A. (1977), ‘The pragmatics of lexical specificity’ *Journal of linguistics*, 13: 153-164.
- 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.
- Franken, N. (1997), ‘Vagueness and approximation in relevance theory’, *Journal of pragmatics*, 28: 135-151.
- Fronek, J. (1982), ‘Thing as a function word’, *Linguistics*, 20: 633-654
- Gleeson, H.A. (1956), *Introductory linguistics*. Philadelphia: Chilton.
- Grice, H.P. (1975), ‘Logic and conversation’, in: P. Cole and J.L. Morgan (eds.) *Syntax and semantics: speech acts*. New York: Academic Press, 41-58.
- Hockett, C.F. (1958), *A course in modern linguistics*. New York: Macmillan.

- Jackson, F. (2002), 'Language, thought and the epistemic theory of vagueness', *Language and communication*, 22: 269-279.
- Jefferson, G. (1990), 'List construction as a task and resource', in: G. Psathas (ed.) *Interaction competence*. Lanham, MD: University Press of America, 63-92.
- Kyburg, A and M. Morreau (2000), 'Fitting words: vague language in context', *Language and Philosophy*, 23: 577-597.
- Louw, B. (1993), 'Irony in the text or insincerity in the writer? the diagnostic potential of semantic prosodies', in: M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology: in honour of John Sinclair*, Amsterdam: Benjamins, 157-176.
- Lybbert, E.K. (1972), 'Substitutes as replacements', *Linguistics*, 91: 5-16.
- Macaulay, R.K.S. (1991), *Locating dialect in discourse: the language of honest men and bonnie lasses in Ayr*. New York: Oxford University Press.
- Manor, R. (1995), 'Pragmatic considerations in semantic analyses', *Pragmatics and cognition*, 3(2): 225-245.
- McCarthy, M.J. and A. O'Keeffe (2002), 'Vague language and participation framework: indices of identity among group, culture and nation' paper read at the 35th Annual Conference of the British Association for Applied Linguistics, University of Cardiff, 12th -14th September, 2002.
- McCarthy, M.J. and A. O'Keeffe (2003), ' "What's in a name?" - vocatives in casual conversations and radio phone in calls', in: P. Leistyna and C. Meyer (eds.) *Corpus analysis: language structure and language use*, Amsterdam: Rodopi, 153-185.
- Mervis, C.B. and E. Rosch (1981), 'Categorization of natural objects'. *Annual review of psychology*, 32: 89-115.
- Morris, P. (ed.) (1994), *The Bakhtin reader*. London: Arnold.
- O'Keeffe, A. (2002), '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, 91-113.
- Oliver, E. (2003), 'No ordinary Joe'. *Irish Times*, February, 15th, page 5.
- Overstreet, M. and G. Yule (1997), 'Locally contingent categorization in discourse', *Discourse processes*, 23: 83-97.
- Pinkal, M. (1985), *Logic and lexicon*. Dordrecht: Kluwer Academic Publishers.
- Powell, M.J. (1985), 'Purposeful vagueness: an evaluative dimension of vague quantifying expressions', *Journal of linguistics*, 21: 31-50.
- 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.
- Rosch, E. (1978), 'Principles of categorization', in: E. Rosch and B. Llyod (eds.) *Cognition and categorization*. Hilldale, N.J.: Erlbaum, 27-48.
- Sinclair, J. McH. (1996), 'The search for units of meaning', *Textus*, IX: 75-106.
- Ward, G. and B. Birner (1992), 'The semantics and pragmatics of "and everything"', *Journal of pragmatics*, 19: 205-214.

- Warren, M. (1993), 'Inexplicitness: a feature of naturalness in conversation', in: M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and technology: in honour of John Sinclair*. Amsterdam: John Benjamins, 37-53.
- Williamson, T. (1994), *Vagueness*. London: Routledge
- Zhang, Q. (1998), 'Fuzziness – vagueness – generality – ambiguity', *Journal of pragmatic*, 29: 13-31.