13   Vagueness

Winnie Cheng and Anne O'Keeffe

13.1   Introduction

Vague language (VL) (Kempson 1977) is a natural and an integral part of everyday discourse (Carter and McCarthy 2006), often viewed as similar to uncertainty (Stubbs 1996), and discussed in relation to hedging, generality, ambiguity, ambivalence, and fuzziness (Chafe 1982, Franken 1997, He 2000, Zhang 1998). VL can also be 'underspecifying', which is common in conversation (Rühlemann 2007: 75). Conversational text need not be self-contained and self-explanatory because conversationalists can rely on rich non-linguistic resources of context (see also Channell 1994, Leech 2000, O'Keeffe, McCarthy and Carter 2007). VL is linguistically manifested in a variety of ways, for example vague additives (including vague approximators and vague tags), vagueness by choice of words and vague quantifiers, vagueness by implicature (Channell 1994), and vague lexis, vague reference (e.g., non-anaphoric pronouns and adverbs and indefinite pronouns) (Cutting 2007).

A notion tightly related to VL is reference. In the classic text on reference, Strawson (1950: 326) distinguishes between reference and denotation and links reference to contextual factors, saying that reference 'is not something an expression does; it is something that someone can use an expression to do'. Crucially, he stresses the salience of 'the context of an utterance' (ibid.: 336). He defines context as 'the time, the place, the situation, the identity of the speaker, the subjects which form the immediate focus of interest, and the personal histories of both the speaker and those he is addressing' (ibid.: 336). Brown and Yule (1983) take Strawson's thoughts on reference further by conceptualising what a 'reference act' on the part of the speaker entails for the speaker–hearer interaction. They note that the concept which interests the discourse analyst is 'successful reference'. 'Successful reference', according to Brown and Yule (1983), 'depends on the hearer's identifying, for the purposes of understanding the current linguistic message, the speaker's intended referent' (ibid.: 205). With this in mind, let us consider what happens when we use vague and underspecified language in successful reference, as shown in the following example, taken from an American talk show.

Example 1
>   Do you need any help, donations, **stuff like that**?
>   ('Talkback_CNN' Corpus of Contemporary American English)

For the vague category marker or tag (Channell 1994) *stuff like that* to be a 'successful reference' (Brown and Yule 1983), the speaker relies even more on the hearer's ability to access the 'rich linguistic resources of context', as referred to by Rühlemann (2007: 75). Brown and Yule (1983) note that when a speaker uses reference to pick out an individual entity, the speaker typically takes into consideration the hearer's developing discourse representation, which the speaker can depend on the hearer being able to use in identifying the intended referent. Schwarz-Friesel and Consten (2011: 351) suggest that hearers 'enrich the text base by incorporating both information from the text and information activated through conceptual instantiation and inferential processing'. O'Keeffe (2004) conducted a detailed empirical analysis of vague category markers, such as *stuff like that*, *things like that*, and *and the like*, in the context of a radio phone-in, and she concluded that both the speaker and the hearers draw on a communal understanding of the boundaries of the shared schematic knowledge which these items refer to. She sees these chunks of language as functioning in an expedient way as linguistic triggers employed by speakers in reference and decoded by participants who draw on their store of shared knowledge. She argues that the meanings of these vagueness markers are socio-culturally grounded and are co-constructed within a social group that has a shared socio-historic reality. Interestingly, Cutting (2000) notes that vague language can fail to communicate when the speaker fails to appropriately assess the hearer's knowledge or state of mind and does not provide enough background information. She also notes that the use of vague language fails to successfully communicate meanings when the speaker misses the reference and implicatures in the case of different cultural backgrounds between the speaker and the hearer. We will return to this point later in the paper in the context of our study.

As we have already mentioned, vagueness is a common and well-cited feature of spoken language in particular. Schwarz-Friesel and Consten (2011: 351) underscore this, saying that referential underspecification is an essential feature of the relation between verbal expressions and the conceptual representations they are intended to express. They remark that 'systematic underspecification is to be regarded as default' and that 'in accordance with some principle of cognitive economy, underspecification is achieved by selecting only the really relevant information and thereby avoiding redundancy' (ibid.: 351).

Regarding the discussion about the levels of specificity, Cruse (1977) explores the notion of unmarked or neutral levels of specificity in various contexts which are not necessarily covered by Grice's (1975) maxims, and 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 underspecification de-emphasises the feature that is omitted, while overspecification emphasises or intensifies the added feature (ibid.: 163). Apart from underspecifying 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.

Vague language takes many forms and in this paper we are particularly interested in vague approximators (Channell 1994), particularly the role of contexts of interaction in speakers' choice of vague approximators and the effect of their linguistic choices on meaning underspecification (Cruse 1977, Rühlemann 2007). Example 2, from the Limerick Corpus of Irish English (LCIE) (Farr, Murphy and O'Keeffe 2004), displays different levels of specificity around the lexical word *money*, achieved by means of co-selection of vague approximators (Channell 1994). To paraphrase the words of Cruse (1975, 1977), both variations of form are equally appropriate from the point of view of the speaker and the listener in a given context of use.

Example 2   (LCIE)
> (a) I saved **some money**.
> (b) I said I would give him **a certain sum of money** . . .

Example 2(a), with *some*, is not very specific. Example 2(b), with *certain*, is an example of purposeful vagueness where, for reasons of discretion, the speaker is not specifying the sum of money but he clearly knows what the amount is.

In this paper, we review the research on vague language, with a focus on the contribution of corpus analysis. The primary aim of the study is to conduct a corpus-based analysis of one aspect of vague language use, namely Channell's (1994) VL type 1 vague approximators, e.g., *about*, when they are used to modify numbers or quantifiers in the context of reference. Specifically, our aim is to illustrate the use of the corpus method to find out what happens to how speakers talk about numbers and quantifiers when they are 'purposefully and unabashedly vague' (Channell 1994). In addition, the study also examines and compares the nouns and noun phrases which are referred to and qualified by vague approximators in the corpora. We classified these nouns and noun phrases into semantic categories in order to take a closer look at vague approximators use in 'successful reference' across two different spoken corpora.

13.2   Vague language types

Channell (1994) divides vagueness into three categories:

(i) vague additives (which include vague approximators (e.g., *about*) and tags referring to vague categories (e.g., *and things like that*),

(ii) vagueness by choice of words (e.g., *yoke*, *thingy*) and vague quantifiers (e.g., *piles of*), and

(iii) 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) (Channell 1994: 18).

After Channell (1994), Cutting (2007) adds other British VL forms and structures that are highly dependent on shared background and interpersonal context for meaning making. They are vague lexis, vague reference, and vague clausal or utterance-level features. Vague lexis includes metonymical proper nouns (as in 'I haven't done any *Chomsky*' and 'Are you going to *Stylistics*?') and superordinate nouns (as in 'You do *Language Planning*, don't you?'). Examples of vague reference are non-anaphoric pronouns and adverbs (as in '*They* sent me *this*' and '*He*'s nearly *there*') and indefinite pronouns (as in 'Did you do *anything* this weekend?' and '*Everybody* was there'). An example of vague clauses is 'Are you going to do *what you thought you'd do?*'; an example of unfinished sentences is 'They had the mental *and the* ...'; and examples of conversational implicature that points to shared knowledge are 'Well *you know what he's like*' and 'I can imagine *why you wouldn't want to*' (Cutting 2007).

13.3   Previous studies of vague language

The study of VL has its origins in the study of semantics, and we are indebted to the antecedent work which has contributed greatly to the overall lineage, in particular the work of Cruse (1975, 1977). VL studies have been conducted in a range of spoken and written discourses and genres, including biomedical slide talks (Dubois 1987), academic writing on economics (Channell 1990), advertising (Myers 1994), telephone conversations (Urbanová 1999), student talk (Jucker, Smith and Ludge 2003), radio phone-ins (O'Keeffe 2004), direct helpline calls and hospital-chaplain–patient interactions (Adolphs, Atkins and Harvey 2007), poetry (Cook 2007), and conference presentations (Trappes-Lomax 2007).

VL research has also been examined and compared in the contexts of different cultures and spoken contexts, for instance, academic, business and public discourses in Hong Kong (Cheng 2007), German and New Zealand English native speaker conversations (Terraschke and Holmes 2007), British and Irish English conversations (Evison, McCarthy and O'Keeffe 2007), English and German conversations (Overstreet 2007), spoken interactions in North American and UK offices (Koester 2007), British courtrooms (Cotterill 2007), and Middle English (Carroll 2009).

Studies concur that VL performs important social, interpersonal and interactional functions by making conversations sound less formal and less direct.

Specifically, the communicative, strategic function of VL for mitigating face threat and maintaining interpersonal relations has been widely discussed. In conversations, for example, VL is used 'to avoid an excess of precision', or to achieve 'imprecision' or 'imprecise language use' (Crystal and Davy 1975: 112–114). VL contributes to 'naturalness and the informal, convergent tenor of everyday talk' (McCarthy 1998: 108–118). It 'nearly always enables polite and non-threatening interaction' (Carter 2003: 92). VL allows 'a speaker to take refuge in strategic imprecision' (Leech 2000: 695). Carter and McCarthy (2006: 202) suggest that VL 'softens expressions so that they do not appear too direct, or unduly authoritative or assertive'. In her study of Hong Kong English, Cheng (2007) notes that speakers' successful use of vagueness indicates a high level of interactivity, particularly in highly context-dependent conversation where responsibility for meaning making is shared among speakers.

Analysing student interactions, Jucker, Smith and Ludge (2003) found that students use VL to perform multiple functions, including to deal with a lack of information, maintain fluency, reduce processing costs, convey a different meaning from that of a precise expression, provide information about the significance of the quantity expressed, convey low levels of certainty, soften implicit complaints and criticisms, and establish a social bond. Smith (2001) finds that teachers' use of VL significantly and adversely affects student achievement scores.

Vagueness can be linguistically realised by means of lexical words or phrases, scalar implicature, and numbers or quantities (Channell 1994; Cutting 2000, 2007). Crystal and Davy (1975: 112–114), for instance, identify types of lexical vagueness, ranging from items which express 'total vagueness', e.g., *thing* and *whatsit*, to examples such as *I've got some tomatoes, beans and things*, and the use of the suffix *-ish* in colloquial English. Brown and Yule (1983: 8–9) point out that spoken language usually contains a lot of 'general, non-specific' vocabulary. Stubbs (1983: 23) observes that speakers in conversations occasionally supply information with such words as *thingummy* and *what-d'you-call-it*, thus leaving the hearer to fill in the specific information. As regards scalar implicature (Gazdar 1979, Levinson 1983, Channell 1994), speakers select from a set, or scale, of items, e.g., *always*, *often*, *usually*, *sometimes*, and *never*, that can be ranked in relation to one another. It is noted by the researchers that often the scale has precise items at each end of the scale and vague items in between, successful interpretation of which depends on the extent to which the hearer and the speaker have a shared understanding of the relative ranking of the items on the scale.

Another area of VL that has received considerable attention, as referred to above in Example 1, is that of 'vague category markers' (O'Keeffe 2004), also referred to as 'vague category identifiers' (Channell 1994), 'general extenders'

(Overstreet and Yule 1997), 'generalised list completers' (Jefferson 1990), 'tags' (Ward and Birner 1993), 'terminal tags' (Dines 1980, Macaulay 1991), and 'extension particles' (Dubois 1993). They all refer to the use of an exemplar followed by a VL tag, for example, *sport, and things like that*; *furniture, and stuff like that*; and *toys and decorations, and the like*. The *ad hoc* manner in which these categories are created has been highlighted by Barsalou (1983) and Overstreet and Yule (1997). The spontaneity of categorisation and the context-dependent nature of the categories themselves is an important feature of their use in real-time face-to-face interaction.

13.4   Vague approximators

Using VL with numbers or quantifiers in order to make approximations (Channell 1994) is another important area of study. The category of numbers or quantities is employed to signal an interpretation as referring to a range rather than a specific quantity (Dubois 1987, Sigurd 1988, Channell 1994, Ruzaitè 2007, Koester 2007, Cheng 2007). Numbers, as a VL type (Channell 1994), are observed to be combined with hedges, e.g., *close to*, *about*, *around*, *on the order of*, and *something like*, in biomedical slide talks to function as 'imprecise' numerical expressions (Dubois 1987: 531), or with approximators (Channell 1994) to function as VL. Sigurd (1988: 243) notes that the numbers used in approximative expressions are typically round numbers, e.g., 10, 20, 50, and 100. Channell's (1994: 114) example of 'We've got *about five or six* of them, but I'll be only going to talk about *three of them* today' illustrates the function of downgrading, with the use of an approximator with numbers to create an impression that the exact number of informants is not important. Ruzaitè (2007: 213) analyses teachers' and students' spoken academic discourse and finds that vague approximators are used 'to shield their claims against possible criticism, avoid categorical claims, observe the politeness principle and save face'. In Rowland's (2007) study of mathematics classrooms, students are expected to use vague hedges, in combination with an approximator and/or numbers, to make predictions and generalisations in problem-solving activities.

13.5   Method of study

As evidenced from the review of the main studies into VL, a number of them are based on corpora. Corpus linguistics offers a lot to the study of VL but it also has certain operational limitations. VL is not part of the automatic tagging system of a corpus so any study of VL will involve the meticulous trawling of general searches to find and count instances and forms of VL. This constraint limits the applicability of CL to the study of VL. Looking at numbers and

quantifiers is possibly one of the more accessible areas, though it does still involve a lot of manual sorting and disambiguating. Despite the limitations, a corpus provides a useful test bed for the analysis of VL in a real context.

In this study, we take one of Channell's (1994) types of VL, approximator + number (n), and investigate it using three corpora. To do this, we used the search items of *about*, *around*, *round*, *approximately*, *or*, *or so*, *at least*, *at most*, *less*, *more*, *under*, *over*, *and over*. These searches had to be disambiguated through manual concordance sorting so as to arrive at only the relevant structures that contain the search items and 'n' and/or 'm' (where 'n' refers to a number and 'm' refers to a multiplier of the number, e.g., *five* (n) *or ten* (m) *minutes*). Following Channell (1994), the HKCSE (conversation) and LCIE were examined for use of:

  (i) approximator + n (i.e., 'about + n', 'around + n', 'round + n', and 'approximately + n'),
 (ii) 'n or ($n + 1$)',
(iii) 'n or m', 'n or so'.

Example 3, taken from the HKCSE (Cheng, Greaves and Warren 2008), illustrates the aspects of VL that are examined in this study.

Example 3   (HKCSE)
(a, Hong Kong Chinese, female; B, native English speaker, male)

> a: What's the price that you bought per per square [per square feet
> B: [four
> B: **about four thousand eight** four thousand nine **nearly four nine**
> a: four nine per square feet you mean
> B: yea

In Example 3, we find the use of vague approximator + number (n) (*about four thousand eight*). On one level, speaker B, by underspecifying the exact cost so as to avoid flouting Grice's (1975) maxims of quantity (specifically 1, 'make your contribution as informative as required', and 2, 'do not make your contribution more informative than is required') but the speaker is also simultaneously truncating the referent (*per square feet*) in real time. The price starts off as *about four thousand eight* or *four thousand nine*, a truncation of 4,800–4,900. This is then reiterated as the even more truncated *four nine* (4,900), preceded by the vague approximator *nearly*. Speaker a, the listener, ratifies this by repeating the most truncated number *four nine*, followed by the noun phrase *square feet* being qualified by the number. Clearly, from Example 3, when speakers choose to underspecify in their reference, or be vague, they are drawing on assumptions and expectations about the 'givenness' of the shared broader social and cultural knowledge as well as the immediate local context of interaction.

Table 13.1 *Breakdown of the Hong Kong and Limerick corpus data used in the study.*

Corpus	Total words	Number of conversations	Percentage of corpus
Hong Kong Corpus of Spoken English sub-corpus: Hong Kong Chinese (HKC)	108,182	71	49.87%
Hong Kong Corpus of Spoken English sub-corpus: Native English speakers (NES)	108,760		50.13%
Limerick Corpus of Irish English (LCIE)	1,000,000	367	100%

13.6   Corpus data

We analyse two spoken language corpora, namely the conversation sub-corpus of the Hong Kong Corpus of Spoken English (HKCSE) (216,942 words). The Hong Kong Corpus of Spoken English (HKCSE) is a one-million word corpus of spoken English collected in Hong Kong (Cheng 2003, Cheng et al. 2008) comprising academic, conversation, business and public sub-corpora. The conversation sub-corpus contains both Hong Kong Chinese (HKC) and native English speakers (NES) in each conversation. The second corpus used in this study is the LCIE, a one-million word corpus of spoken English collected in Ireland (Farr, Murphy and O'Keeffe 2004). We draw on these two spoken corpora from different varieties of English in order to explore sociocultural dimensions to VL use. The conversational data used in this corpus-based study are described in Table 13.1.

The dataset provides a very interesting opportunity for comparison. On the one hand, it allows for intracultural comparison by comparing results from the HKC with those from the NES, whereby the conversations between native and non-native English speakers are compared. In each of the 71 conversations in the HKCSE, there are both Hong Kong Chinese (HKC) and native English speakers (NES). On the other, simultaneously, we have the opportunity for intercultural comparison of the results from Hong Kong English with those of Irish English. Therefore, this analysis is both an intracultural and an intercultural study. The entire LCIE was used in the analysis of vague approximator + number (n). In order to make the LCIE results comparable, they have been normalised to occurrences per 10,000 words.

13.7   Findings and discussion

Table 13.2 summarises the findings (raw frequency and normalised (N) per 10,000 words) between the HKC and NES in the HKCSE (conversation), for each of the thirteen sub-types of vague approximator + number (n).

Table 13.2 *HKC versus NES in HKCSE (conversation): comparison of thirteen sub-types of VL type 1.*

Type 1 VL sub-types	Hong Kong Chinese (HKC)		Native English Speaker (NES)		Total HKCSE (conversation)	
	Freq.	N	Freq.	N	Freq.	N
about $+ n$	83	7.67	90	8.28	173	7.97
around $+ n$	7	0.65	1	0.09	8	0.37
round $+ n$	1	0.09	0	0	1	0.05
approximately $+ n$	0	0	0	0	0	0.00
n or $(n + 1)$	34	3.14	47	4.32	81	3.73
n or m	17	1.57	13	1.20	30	1.38
n or so	2	0.18	1	0.09	3	0.14
at least $+ n$	16	1.48	6	0.55	22	1.01
at most $+ n$	0	0	0	0	0	0.00
less $+ n$	7	0.65	3	0.28	10	0.46
more $+ n$	15	1.39	5	0.46	20	0.92
under $+ n$	3	0.28	1	0.09	4	0.18
over $+ n$	7	0.65	12	1.10	19	0.88
Total	192	17.75	180	16.55	372	17.15

A chi-squared test of the results tells us that none of the differences between the two corpora are significant. The findings here show that a total of 372 instances of type 1 VL are used by the speakers in the HKCSE recorded in Hong Kong. Among the 372 instances, the most frequent five sub-types are 'about $+ n$' ($N = 174$), 'n or $(n + 1)$' ($N = 81$), 'n or m' ($N = 30$), 'at least $+ n$' ($N = 22$), and 'more $+ n$' ($N = 20$). The HKC speakers are found to use type 1 VL slightly more frequently than the NES (17.75 compared with 16.55 occurrences per 10,000 words). But again, the difference is hardly noteworthy, given its insignificance ($p = 0.79$). Among the thirteen sub-types of type I VL, the two groups of speakers do not use any 'approximately $+ n$' and 'at most $+ n$'. For the remaining 11 sub-types, the HKC use a larger proportion of VL in eight of them, except for 'about $+ n$', 'n or $(n + 1)$', and 'over $+ n$'. Ostensibly, the greatest usage differences are found in 'n or $(n + 1)$', where NES exceeds HKC by 1.18, and 'at least $+ n$', where HKC exceed NES by 0.93 but again this is not statistically significant ($p = 0.99$ for each).

Table 13.3 and Figure 13.1 compare the results for the Irish English data from LCIE with those of the HKC and NES in the HKCSE (as presented in Table 13.3), across the thirteen sub-types of type 1 VL. The normalised results for thirteen sub-types of type 1 VL show no significant differences in frequency, though there are marginally fewer instances of use in total in the LCIE (12.27)

Table 13.3 *LCIE and HKCSE (conversation): comparison of thirteen sub-types of VL type 1.*

| Type 1 VL sub-types	LCIE		HKCSE (HKC)		HKCSE (NES)		Total HKCSE (conversation)	
	Freq.	N	Freq.	N	Freq.	N	Freq.	N
about + *n*	930	9.3	83	7.67	90	8.28	173	7.97
around + *n*	112	1.12	7	0.65	1	0.09	8	0.37
round + *n*	0	0	1	0.09	0	0	1	0.05
approximately + *n*	6	0.06	0	0	0	0	0	0.00
n or (*n* + 1)	281	2.81	34	3.14	47	4.32	81	3.73
n or *m*	122	1.22	17	1.57	13	1.20	30	1.38
n or so	27	0.27	2	0.18	1	0.09	3	0.14
at least + *n*	9	0.09	16	1.48	6	0.55	22	1.01
at most + *n*	0	0	0	0	0	0	0	0.00
less + *n*	9	0.09	7	0.65	3	0.28	10	0.46
more + *n*	27	0.27	15	1.39	5	0.46	20	0.92
under + *n*	29	0.29	3	0.28	1	0.09	4	0.18
over + *n*	44	0.44	7	0.65	12	1.10	19	0.88
Total	1,227	12.27	192	17.75	180	16.55	372	17.15
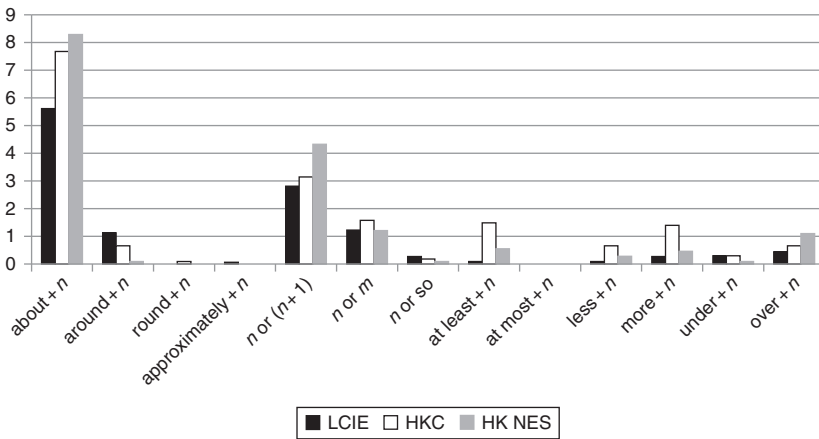
Figure 13.1 HKCSE (conversation) and LCIE: comparison of thirteen VL type 1 sub-types.

than in the HKCSE (17.15). The results of the chi-squared test confirm that they are not statistically significant. On the surface, therefore, the comparative results suggest a degree of universality in form choice when using a vague reference in approximating with numbers. In order to delve deeper into the data, there is a need to look more closely at how the speakers use the VL forms. To this end, we will now focus on the most frequent approximator sub-type 'about + *n*' in the corpora to investigate how it is used. This will bring us closer to how the approximator is being used in reference and in particular in terms of what is being referred to. Similar to the HKCSE data, the sub-type 'about + *n*' in the LCIE is the most frequent (9.3), which is slightly greater than the HKCSE data (7.97 in total). Extracts of 'about + *n*' from the LCIE are presented, as follows.

Example 4   'about + *n*' (LCIE)
        Remember the last time before you went to Spain there was **about fifteen** of us in the house.
        . . . there was a big party **about two weeks** ago for her twenty-first.

In the LCIE, the sub-type '*n* or (*n* + 1)' is the second most frequent item (2.81) although its frequency is lower than that of either speaker group of the HKCSE (HKC = 3.14; NES = 4.32). Example 5 illustrates '*n* or (*n* + 1)'.

Example 5   '*n* or (*n* + 1)' (LCIE)

        A: And has Roisin any notion of settling down?
        B: Why? Is she going with someone?
        A: For years. **About eight or nine** years.

While the sub-types 'about + *n*' and '*n* or (*n* + 1)' are most frequent in all of the corpora, correspondingly, 'round + *n*', 'approximately + *n*', '*n* or so', and 'at most + *n*' are the lowest or non-existent in all the corpora.

A qualitative analysis of the HKCSE, in terms of what is being referred to in the form of the linguistic realisations of the nouns/noun phrases modified by vague approximator + number (*n*), shows that vagueness is expressed in similar semantic categories and with a comparable range of categories. Across all the thirteen sub-types and between the HKC and the NES, the most frequent semantic categories (underlined), in descending order of frequency, are 'calendar period' (*day*(*s*), *week*(*s*), *month*(*s*), *year*(*s*)), 'not identified', 'concrete and abstract objects', 'time' (*minute*(*s*), *hour*(*s*)), 'unit of money', 'frequency', 'ratio', 'distance/length', and 'vague referents' (*thing*).

The following are extracts from the Hong Kong conversations to provide contextual examples for both the HKC and the NES.

Example 6   (HKC in HKCSE)
        I er I er lived in Canada for **about six years** and I studied there.
        . . . hospital er only promise to give a treadmill **about eight months** afterwards.

. . . you know I think **about four o'clock** something.
. . . it's a long I think it's **about eight kilometres**.
I think at **about thirty-five per cent**.
. . . we have er **about twenty offices** overseas with sales team . . .
. . . we have hubs all over the world **about five places** in London as well.

Example 7   (NES in HKCSE)
. . . is that **about twenty years** ago and so . . .
. . . he was doing a course at the university for **about nine months**.
. . . started putting it onto the computer till Friday at **about eleven o'clock**.
. . . it's quite long actually it's **about four feet** long.
. . . probably **about ninety-five per cent** back.
. . . he needs to take **about twenty pills** every morning.
I usually get them for **about sixteen bucks**.
I think it was er near **about twenty two degrees**.

Below, all instances of 'about + *n*' in the HKC, NEC, and LCIE are analysed in terms of the semantic categories used with 'about + *n*' (HKC, 83; NES, 93; LCIE, 930). A random sample of 100 occurrences was analysed for each dataset. Figure 13.2 compares the referents of 'about + *n*' in the corpora, in occurrences per hundred, and shows that time or calendar periods are by far the
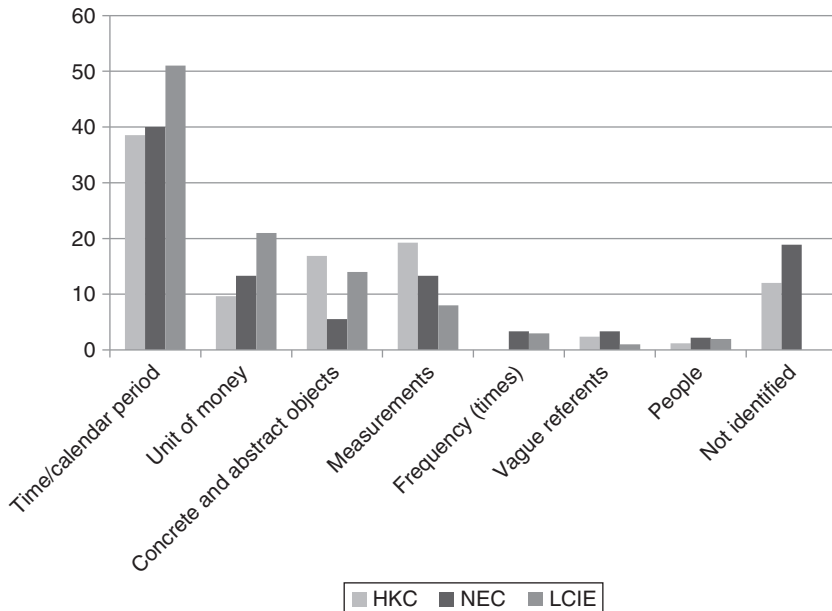

Figure 13.2 Breakdown of semantic categories used with 'about + *n*', per 100 occurrences in HKC, NEC, and LCIE.

most frequent semantic category referred to when using the approximator *about* (see Examples 8–10).

Example 8 'about + *n*' (HKC in HKCSE)
  I haven't been right down there I think **about four <u>years</u>** something.
  ... and every lesson it takes **about half <u>hour</u> half an <u>hour</u>** ...
  You have to spare **about one hour thirty <u>minutes</u>**.
  I went um to the UK for **about three <u>weeks</u>** just to visit my friends there.
  ... when we have the first baby and that was er **about nine <u>year</u>**s ago and we've used er four maids ...

Example 9 'about + *n*' (NES in HKCSE)
  ... it's too long and then I hope it it lasted **about forty <u>minutes</u>** but whatever.
  ... and um open presents and then go to sleep **about one <u>o'clock</u>** wake up and then ...
  I've been to Pattaya **about six <u>years</u>** ago ...
  I think she needs **about twenty <u>hours</u>** which doesn't sound very much.

Example 10 'about + *n*' (LCIE)
  [referring to someone who appeared on television] He was only on **about five <u>minutes</u>** like.
  She said go in now after **about five <u>minutes</u>** so I walked ...
  [referring to a computer virus] It knocks off everything after **about a <u>minute</u>**.
  +ah **about two <u>weeks</u>** ago Triona went out one night ...
  About six <u>hours</u> in all.

A further analysis of the concordance lines of 'about + *n*' shows some points of note. First, we notice culturally implicit VL. In the Hong Kong extract below (Example 10), speaker b (HKC, male) explains the size of the domestic helper's bedroom to speaker B (NES, male). In addition to the use of *about*, the negotiation of meaning is concerned with the measurement unit of 'square feet'; and implicitly, how it compares with the metric system of 'metre' and 'square metre'.

Example 11 (HKCSE)

    B: mm [so did she have a room in   [your house
    b: [er   [oh yes yes yes yes she has a room there well
             actually there's is a maid's room in in [in the back of the quarter
    B: [how big would that be
    b: sorry
    B: how big would it be
    b: well (.) the room I think it's **about** er
         (pause)
    b: **sixty to seventy <u>square feet</u>** I think

Another point of interest concerns the relationship of VL and directness. While VL is most often associated with face-saving (Carter and McCarthy 2006), we found in the context of family discourse that it can be open to challenge in a very direct way. Directness in family discourse is noted by Clancy (2005), who points out that because of the stable or fixed nature of kinship, much more directness is afforded within family discourse (behind closed doors). In Example 11, we see an interaction between a mother and a son. We also find use of 'about + *n*' where the vagueness is challenged in this mother–son interaction in a seemingly very direct and face-threatening interaction, yet it ends with a normal leave-taking when the mother says *Okay bye*.

Example 12 (LCIE)
(A, the son; B, the mother)

> A: I'll be back later.
> B: Okay.
> A: Good luck.
> B: What time is later Jason?
> A: **About half an hour**.
> B: **Half an hour?**
> A: I don't know. I'd say an hour probably.
> B: Are you definitely going to be back then?
> A: Why?
> B: Cause I'll lock the door if you aren't.
> A: Lock the door away.
> B: Okay.
> A: I'll open it myself.
> B: Okay bye.
>   <sound of car driving off>

In our analysis so far, we have seen that on the surface, reference as manifested through vague approximators seems to be a universal feature in both form and distribution with no significant quantitative differences emerging either between the intravarietal context in the Hong Kong data or intervarietally in relation to the Irish data and the Hong Kong data. However, when we delved deeper into the actual reference processes, we found cultural implicitness and also contextually specific uses of the approximator *about* + number (e.g., in the use of directness in family data).

   Let us return to another aim of this paper, namely the examination of VL using a corpus. As we noted, it is a challenging task to analyse VL as a whole in corpus data simply because corpora are not annotated for VL, unless in the case of small specialised corpora. This means that manual reading of texts is required for any large scale empirical study of VL. What we hope to have shown, however, is that aspects of VL can be studied successfully using corpora and these are a good starting point for anyone interested in the study

of aspects of VL in a corpus (see O'Keeffe 2004, Cheng 2007, Terraschke and Holmes 2007, Evison, McCarthy and O'Keeffe 2007, Koester 2007, Cotterill 2007, to name a few). We also hope to have illustrated the usefulness of taking Channell's (1994) VL taxonomy and working with its syntactic parameters to search through corpora for instances of use of VL. While it did involve a lot of manual sorting through concordance lines to eliminate non-VL instances, it was not by any means an insurmountable task. However, we do note the limitation that the set of features on which we focused our search had to be predefined. Nonetheless, this does allow for others to make comparisons using the same parameters, but with different datasets.

By applying this framework, we have compared spoken data from culturally different sources. Having a common framework of VL analysis means that this study can be easily replicated using other corpora. What is interesting is that we did not find stark differences in the distribution of VL linguistic forms in our case study of Hong Kong Chinese and native English speaker interactions and Irish English interactions. Though varietally different, the corpora show the same profile in terms of the most frequent and the least frequently used forms. When we examined the most common form (i.e., 'about $+ n$') in greater detail across the corpora, we found that the profile of what it qualifies shows a lot of similarity of distribution, with VL around time and calendar periods being the most frequent. In any corpus study, the numbers must be interrogated qualitatively and our close examination of VL used with time and calendar periods has revealed some culturally specific uses of vagueness as well as contextually specific uses.

Although the frequency and analysis of the semantic categories qualified by vague approximators $+$ numbers display no statistically significant differences in terms of the distribution of the sub-types of vague approximators $+$ numbers across the corpora, the corpus examples demonstrate the effectiveness in meaning making in different interpersonal and cultural contexts.

Returning to the notion of reference set out at the beginning of this paper, where we cited the notion of 'successful reference' (Brown and Yule 1983: 205) and its dependence on 'the hearer's identifying, for the purposes of understanding the current linguistic message, the speaker's intended referent', Example 11 illustrates very nicely a case where 'successful reference' has to be worked at: the use of VL in the communication between a Hong Kong Chinese and a native English speaker required further explanation before the meaning of the referent could be effectively conveyed to the hearer. The HKC describes the size of the room with *about er sixty to seventy square feet*. The NES then requests an elaboration of how big it actually is. Such a question or request reflects the hearer's failure in understanding the size of the room. A possible reason for such communication failure could be due to the speaker's failure in appropriately assessing the hearer's knowledge or background understanding

(Cutting 2000). The making and interpretation of meanings are the responsibility of both the speaker and the hearer (Cheng 2007). In this case, perhaps the unit of measurement is not shared knowledge between the speaker and the hearer. With reference to Cruse's (1977) notion of 'givenness', this example illustrates a discrepancy in the 'givenness' between the speakers' expected and actual social and cultural knowledge.

In spite of the amount of time spent on manually sorting the concordance lines of VL and non-VL instances, the larger contexts extracted from the concordance lines, i.e., corpus examples, are useful in providing more information of the communication with VL. Such authentic language data demonstrate not only the actual linguistic realisations of VL, but also the contexts in which they are adopted, and the contribution or hindrance they might have in impacting the effectiveness of communication between interlocutors.

13.8   Conclusion

Using a type of VL from Channell's (1994) framework, this corpus-based study is both intracultural and intercultural, by comparing the use of VL between native and non-native speakers, and by comparing Hong Kong English and Irish English.

Both quantitative and qualitative findings of the present study of vague approximators conclude that the use of VL is closely associated with the notions of 'givenness' (Cruse 1977) and 'successful reference' (Brown and Yule 1983) in the context of VL, which is dependent on an assumption and expectation by the speaker of a high degree of shared social and cultural background knowledge over and above the immediate physical context of the interaction. The corpus examples used in this paper not only demonstrate successful reference in the use of 'purposefully and unabashedly vague' (Kempson 1977; Channell 1994) and underspecified language, but also less successful reference using VL, which may have been caused by the incorrect assumption and expectation of the givenness of the vague approximators + numbers on the interlocutor's part. This brings us back to the classic text on reference, Strawson (1950), which predates corpus linguistics but which pre-empts a crucial point which CL frequently illustrates empirically:

the basic one-way version of discourse communication is quite obviously an abstraction away from the complex interaction which actually takes place between speakers' versions of hearers' versions of speakers' versions (and so on) of representations, in normal discourse situations … At best, the hearer is likely to arrive at a representation which is only partially similar to the speaker's and which, moreover, can only ever be a partial reflection of the so-called 'actual' state of affairs which existed in the world. (Strawson 1950: 206–207)

The study has succeeded in illustrating that corpus analysis provides language researchers with a data-informed understanding of patterns and contexts of language use even in a case where the search items are not normally tagged. It also shows the contribution of corpus research to an important area of study of pragmatics, namely VL use in the context of reference. Despite the operational limitations of a corpus-based study, the analysis of the frequency, forms, and functions of VL, with the use of real world communicative data, allows fuller and more specific descriptions of language.

REFERENCES

Adolphs, S., Atkins, S. and Harvey, K. 2007. Between professional requirements and interpersonal needs: Vague language in healthcare. In J. Cutting (ed.), *Vague Language Explored*, pp. 62–78. Basingstoke and New York: Palgrave Macmillan.

Barsalou, L. 1983. Ad hoc categories. *Memory and Cognition* 11: 211–277.

Brown, G. and Yule, G. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.

Carroll, R. 2009. Vague language in the medieval recipes of the *Forme of Cury*. In M. Peikola, J. Skaffari and S. K. Tanskanen (eds), *Instructional Writing in English: Studies in Honour of Risto Hiltunen*, pp. 55–82. Amsterdam and Philadelphia: John Benjamins.

Carter, R. 2003. Key concepts in ELT: Language awareness. *ELT Journal* 57 (1): 64–65.

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

Chafe, W. 1982. Integration and involvement in speaking, writing, and oral literature. In D. Tannen (ed.), *Spoken and Written Language: Exploring Orality and Literacy*, pp. 35–53. Norwood, NJ: Ablex.

Channell, J. 1990. Precise and vague quantities in writing on economics. In W. Nash (ed.), *The Writing Scholar: Studies in Academic Discourse*, pp. 95–117. London: Sage.

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

Cheng, W. 2003. Humor in intercultural conversation. *Semiotica* 146 (1/4): 287–306.
     2007. The use of vague language across spoken genres in an intercultural corpus. In J. Cutting (ed.), *Vague Language Explored*, pp. 161–181. Basingstoke and New York: Palgrave Macmillan.

Cheng, W., Greaves, C. and Warren, M. 2008. *A Corpus-Driven Study of Discourse Intonation: The Hong Kong Corpus of Spoken English (Prosodic)*. Amsterdam and Philadelphia: John Benjamins.

Clancy, B. 2005. *You're fat. You'll eat them all*. Politeness strategies in family discourse. In K. P. Schneider and A. Barron (eds), *The Pragmatics of Irish English*, pp. 177–199. Berlin: Mouton de Gruyter.

Cook, G. 2007. 'This we have done': The different vagueness of poetry and public relations. In J. Cutting (ed.), *Vague Language Explored*, pp. 21–39. Basingstoke and New York: Palgrave Macmillan.

Cotterill, J. 2007. 'I think he was kind of shouting or something': Uses and abuses of vagueness in the British courtroom. In J. Cutting (ed.), *Vague Language Explored*, pp. 97–114. Basingstoke and New York: Palgrave Macmillan.

Cruse, D. A. 1975. Hyponymy and lexical hierarchies. *Archivum Linguisticum* VI:
    26–31.
    1977. The pragmatics of lexical specificity. *Journal of Linguistics* 13: 153–164.
Crystal, D. and Davy, D. 1975. *Advanced Conversational English*. London: Longman.
Cutting, J. 2000. *Analysing the Language of Discourse Communities*. Oxford: Elsevier.
    (ed.) 2007. *Vague Language Explored*. Basingstoke and New York: Palgrave
    Macmillan.
Dines, E. 1980. Variation in discourse – and stuff like that. *Language in Society* 1:
    13–31.
Dubois, B. L. 1987. 'Something in the order of around forty to forty-four':
    Imprecise numerical expressions in biomedical slide talks. *Language in Society*
    16: 527–541.
Dubois, S. 1993. Extension particles, etc. *Language Variation and Change* 4:
    179–203.
Evison, J., McCarthy, M. J. and O'Keeffe, A. 2007. Looking out for love and all the rest
    of it: Vague category markers as shared social space. In J. Cutting (ed.), *Vague
    Language Explored*, pp. 138–157. Basingstoke and New York: Palgrave
    Macmillan.
Farr, F., Murphy, B. and O'Keeffe, A. 2004. The Limerick Corpus of Irish English:
    Design, description and application. *Teanga (Yearbook of the Irish Association for
    Applied Linguistics)* 21: 5–29.
Franken, N. 1997. Vagueness and approximation in relevance theory. *Journal of
    Pragmatics* 28 (2): 135–151.
Gazdar, G. 1979. *Pragmatics: Implicature, Presupposition and Logical Form*. New
    York: Academic Press.
Grice, H. P. 1975. Logic and conversation. In P. Cole and J. L. Morgan (eds), *Syntax
    and Semantics: Speech Acts*, pp. 41–58. New York: Academic Press.
He, Z. R. 2000. A further study on pragmatic vagueness. *Journal of Foreign Languages*
    1: 7–13.
Jefferson, G. 1990. List construction as a task and resource. In G. Psathas (ed.),
    Interaction Competence, pp. 63–92. Lanham, MD: University Press of America.
Jucker, A. H., Smith, S. W. and Lüdge, T. 2003. Interactive aspects of vagueness in
    conversation. *Journal of Pragmatics* 35: 1737–1769.
Kempson, R. M. 1977. *Semantic Theory*. Cambridge: Cambridge University Press.
Koester, A. 2007. 'About twelve thousand or so': Vagueness in North American and
    UK offices. In J. Cutting (ed.), *Vague Language Explored*, pp. 40–61. Basingstoke
    and New York: Palgrave Macmillan.
Leech, G. 2000. Grammars of spoken English: New outcomes of corpus-oriented
    research. *Language Learning* 50 (4): 675–724.
Levinson, S. 1983. *Pragmatics*. Cambridge: Cambridge University Press.
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.
Myers, G. 1994. *Words in Ads*. London: Arnold.
O'Keeffe, A. 2004. 'Like the wise virgins and all that jazz': Using a corpus to examine
    vague language and shared knowledge. *Language and Computers* 52 (1): 1–20.

O'Keeffe, A., McCarthy, M. J. and R. A. Carter 2007. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press.

Overstreet, M. 2007. Pragmatic expressions in cross-linguistic perspective. *Applied Research in English* 1 (2): 1–13.

Overstreet, M. and Yule, G. 1997. Locally contingent categorization in discourse. *Discourse Processes* 23: 83–97.

Rowland, T. 2007. Well maybe not exactly but it's around fifty basically? Vague language in mathematics classrooms. In J. Cutting (ed.), *Vague Language Explored*, pp. 79–96. Basingstoke and New York: Palgrave Macmillan.

Rühlemann, C. 2007. *Conversation in Context: A Corpus-Driven Approach*. London: Continuum.

Ruzaitè, J. 2007. Vague references to quantities as a face-saving strategy in teacher–student interaction. *Lodz Papers in Pragmatics* 3: 157–178.

Schwarz-Friesel, M. and Consten, M. 2011. Reference and anaphora. In W. Bublitz and N. R. Norrick (eds), *Foundations of Pragmatics*, vol. 1 of Handbooks of Pragmatics, pp. 347–372. Berlin: Mouton de Gruyter.

Sigurd, B. 1988. Round numbers. *Language in Society* 17: 243–252.

Smith, B. 2001. Fiat objects. *Topoi* 20 (2): 131–148.

Strawson, P. F. 1950. On referring. *Mind* 59: 320–344.

Stubbs, M. 1983. *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. Oxford: Blackwell.

  1996. *Text and Corpus Analysis*. Oxford: Blackwell.

Terraschke, A. and Holmes, J. 2007. 'Und Tralala': Vagueness and general extenders in German and New Zealand English. In J. Cutting (ed.), *Vague Language Explored*, pp. 198–220. Basingstoke and New York: Palgrave Macmillan.

Trappes-Lomax, H. 2007. Vague language as a means of self-protective avoidance: Tension management in conference talks. In J. Cutting (ed.), *Vague Language Explored*, pp. 117–137. Basingstoke and New York: Palgrave Macmillan.

Urbanová, L. 1999. On vagueness in authentic English conversation. *Brno Studies in English* 25: 99–107.

Ward, G. and Birner, B. 1993. The semantics and pragmatics of *and everything*. *Journal of Pragmatics* 19: 205–214.

Zhang, Q. 1998. Fuzziness – vagueness – generality – ambiguity. *Journal of Pragmatics* 29: 13–31.