Vaguely speaking: a corpus approach to vague language in intercultural conversations

Neil Drave
City University of Hong Kong

Abstract
This paper presents findings from the initial stage of an ongoing, corpus-based study into the forms and functions of vague language in intercultural conversations. The corpus consists of 30 hours of naturally occurring conversation in English between native speakers of English (NSE) and native speakers of Cantonese (NSC), recorded in Hong Kong. A concordancing program was used to identify all examples of the three main types of vague language and to detect major distributional patterns in a sample of the most frequent items. It was found that, while NSE participants use more vague language than NSC participants, the range of different types was similar for the two groups of speakers, as were major collocations. However, there were important functional differences, notably the greater NSE exploitation of vague language for affective and interpersonal purposes. These are illustrated using the example of stuff.

1. Introduction
This paper reports on ongoing research into the functions of vague language in a corpus of intercultural conversations in Hong Kong. The research derives from an interest in two areas, vague language and intercultural pragmatics, and seeks to address gaps in both areas of research and to find ways in which the study of one can illuminate the study of the other. The study is corpus-based and draws upon familiar work in Gricean pragmatics and Conversation Analysis to inform the explanation of usage differences between native speakers of Cantonese and of English as they converse in English.

In this paper, I first define vague language and briefly summarise relevant research. I go on to describe the corpus of conversations used in the study and present some preliminary findings from the quantitative phase in which a small selection of exemplars were studied using concordancing software. I then look more closely at one vague language exemplar and conclude with some brief remarks on the relevance of corpus methods for researching conversational pragmatics.

2. Vague language
Vague language (hereafter VL) may be defined as that which modifies a linguistic item, phrase or utterance to make its meaning less precise (cf. Channell 1994: 20). In this study VL comprises a closed set of items which are inherently or purposively (Powell 1985: 31) vague in that it is impossible to paraphrase their meaning in a precise (non-vague) fashion. Various terms have been used in the literature to refer to VL, with one influential early study calling it ‘approximation’ (Prince et al. 1982: 85).
Vagueness as defined here has a certain degree of semantic overlap with hedging, since both are concerned with conveying the relationship between a speaker/writer and the matter of an utterance/statement, which places both within the domain of epistemic modality (Hyland 1998: 44). For our present purposes, it may be convenient to think of VL as a sub-category of hedging (cf. Hyland 1998) since VL is more specifically concerned with propositional precision or exactness.

It is necessary to make a distinction between VL, a concern of linguists, and vagueness, a concern of philosophers, and in so doing to work with a natural-language conception of what is vague rather than a logical one. The basic framework for types of VL has been adopted with minor modifications from the seminal work of Channell (1983, 1985, 1990, 1994) and includes the following types of VL:

- **Number approximations**
  A word or phrase is added to a precise figure to signal a vague reading e.g. *about 12, 5 or 6*
  A sub-category of ‘partial specifiers’ indicates that the approximation is either more or less than the named figure e.g. *at least 20, under 100*

- **Non-numerical vague quantifiers (indeterminate)**
  Vagueness is indicated by vague words or phrases rather than by using numbers e.g. *several people, a lot of books*

- **Vague category markers**
  A vague ‘tag’ is added to a lexical item (usually a noun) in order to indicate that this item exemplifies a category (natural or otherwise) e.g. *fish or something, lectures and so on*

- **Placeholder words** e.g. *thing, stuff*
  These words stand on their own as replacements for other, more specific nouns.

Different researchers favour different terms for these VL types. For example, Overstreet (1999) calls category markers ‘general extenders’ because they are non-specific and function to extend otherwise grammatically complete utterances. Overstreet also emphasises the interpersonal function of such items over the categorisation function (*contra* Channell 1994). While I too adopt this kind of approach, for various reasons I have chosen to retain Channell’s terminology.

While these types of VL share the property of propositional imprecision (this is their ‘family resemblance’ in Wittgensteinian terms) it is quite possible for the same VL items to perform different specific functions and for similar functions to be instantiated using different lexis. I therefore follow Hyland (1998: 186) and others in viewing the typology of VL expressions as consisting of several fuzzy categories with protean boundaries.

The major function of VL is to tailor conversational contributions to the perceived informational needs of the other participant(s) so as to maintain and enhance the ongoing relationship. Specific functions include:

- **Filling lexical gaps** (where a speaker cannot recall a word or where one does not exist in the language)
- **Filling knowledge gaps** (memory lapse)
Emphasising (and de-emphasising) certain information  
Deliberately withholding specific information  
Conveying tentativeness  
Conveying an evaluation of, or expectation about, a proposition  
Maintaining an atmosphere of friendliness, informality or deference.

There are epistemic (propositional) and affective meta-functions here: changing the degree of specificity of discourse may reflect knowledge or vocabulary limitations on the part of the speaker, but also contributes to the overall tenor of the conversation and thus to the achievement and maintenance of interpersonal goals. It is very difficult to determine which of the functions (or motives) is in evidence for any single VL item, but this list of functions is nevertheless a necessary, preliminary heuristic for approaching the study of VL.

Previous studies, from which the preceding frameworks were derived, have provided a useful start in describing the scope of reference of VL items in L1 conversation, and situating them broadly within a Gricean pragmatic framework of information manipulation (Grice 1975). However there is much more to be said about where and why VL is used; in other words, about the strategic uses of VL in naturally occurring conversations, for promoting politeness and intersubjectivity (cf. Drave, Cheng and Warren 1996; Overstreet 1999) and for managing asymmetries of knowledge (Drew 1991: 21), particularly in intercultural interaction. In the following section, I describe the aims and methods of the study as a whole, before moving on (in section 4) to talk about the preliminary quantificational findings which are the main focus of this paper.

3. Aims of the research

The central concern of this research is to determine whether intercultural conversations in Hong Kong exhibit vagueness, to discover the nature of vagueness exhibited, and to determine its effect on communication. Intercultural conversations are here restricted to conversations in English between adult native speakers of English (hereafter NSE²) and adult native speakers of Cantonese (NSC), recorded in Hong Kong. The approach is mainly lexical but any implications for discourse coherence will also be considered at a later stage.

While there has been a great deal of work on intercultural interaction (e.g. Piirainen-Marsh 1995), very few studies qualify as useful and defensible discussions of naturally occurring casual conversation. If there is a corpus at all, rarely is the corpus large enough to shed light on interesting linguistic patterns and often the data are so unnatural as to be of limited use for drawing conclusions about how real people behave in the real world (because they are role-plays, for example). This study attempts to rectify these inadequacies using a large corpus of naturally occurring data, as described in the following section.
4. **Corpus**

The corpus for this study was selected from the Hong Kong Corpus of Conversational English (HKCCE), a growing body of intercultural conversational data held at the HK Polytechnic University (for details, see Cheng and Warren 1999). The recordings were done by members of the research team (including myself) and research assistants. The majority selected for my corpus were recorded in 1997 and 1998 in food courts and restaurants (i.e. they are private conversations in public places) using digital mini-disc recorders. Participants’ permission was obtained before the recording began, in line with the guidelines on ethical research promulgated by the HKPU and City University of Hong Kong, and personal details were collected using a short pro-forma.

The details of the corpus are:

- Number of words: 182,518 (NSE 98,310, NSC 84,208)
- Hours: approximately 30
- No. speakers: 222 (NSE 104, NSC 118)
- No. conversations: 84
- Transcribed (using a modified version of Jefferson’s well known conventions)

This is a small corpus compared to many published ones, but for a project of this size it would seem to be large enough. It is unparsed and untagged, except that when using the concordancing program *Wordsmith Tools*, I have identified whether each VL item was spoken by an NSE or NSC.

The conversations in the corpus for this study are considered to be intercultural in the sense that at least one participant was born in Hong Kong and is a native speaker of Cantonese, and at least one other was not born there and is a native speaker of English. Prevailing opinion seems to be that any general notion of ‘culture’ is of limited use when it comes to explaining the moment-by-moment behaviour of participants in interaction (Sarangi 1994, 1995; Shea 1994; Scollon 2001): in other words, even if culture can be identified as a variable in interaction, how is it operationalised and how can one identify its effects? It would seem necessary to adopt an open-minded and socio-critical approach to conversations in which members of different language groups come together and to be wary of ascribing all differences in language to cultural differences, even when these can be relatively easily identified. This ‘non-essentialist and action-oriented’ perspective on culture (Sarangi 1994: 416) is a necessary antidote to a neo-Whorfian (Gumperzian) one, which claims that differences in linguistic behaviour are always due to differences in cultural practices. The choice of the term intercultural should therefore be regarded as a convenience rather than as programmatic, intended to indicate only that there are two groups of speakers with different first languages.

5. **Quantitative findings**

This section presents illustrative preliminary findings from what might be called the piloting phase of the wider research, in which concordancing methods were tried on a
small sample of data. The aims of this phase were to compare the two sets of speakers in terms of the amount and types of VL they used, as well as to try out concordancing methods on a small sample of data to determine whether to apply them to the whole data set. Due to space limitations, only four kinds of findings are reported here:

(i) Amount of VL used by NSE and NSC
(ii) Types of VL used by NSE and NSC
(iii) Patterns: collocation and clustering
(iv) Functional differences between NSE and NSC: the case of *stuff*

### 5.1 Amount of VL used by NSE and NSC

It seems that NSE are ‘vaguer’ than NSC. A frequency count of the number of VL expressions (tokens) which occur more than once revealed that NSE used approximately 1500 in total (59%), while NSC used around 1050 (41%). The total number of lexical items which form part of vague expressions was also counted: in number approximations, for example, the number itself was discounted so that *about 50* was counted as one word while *50 more or less* was counted as three words. It was found that approximately 3.4% of the NSE’s word total is vague, compared to around 2.7% for NSC. The range of vague expressions (types) was similar, however, with only two used exclusively by NSE.

This means that around 3% of the words in the corpus form part of vague expressions, a relatively small proportion compared to similar research but understandable given the narrower scope of the present work. In Nikula’s (1996) study of pragmatic force modification in Finnish-English conversations, 9.7% of words were found to have a modifying function, but the range of devices at issue was much larger in her study since it included modal verbs and other hedges. Kennedy (1987) found that around 14% of words perform a quantifying function, but this figure includes non-vague quantifiers.

### 5.2 Types of VL favoured by NSE and NSC

It was expected that NSE would use VL which NSC did not. The findings were contrary to expectations, with very few VL items being used exclusively by one group or the other and the rank orders of the most frequent items virtually identical. Neither NSE nor NSC use the full range of VL items available in English: of the 150-odd possible VL items, roughly half did not occur at all and seven occurred only once. As an example, a comparison of the most frequently used VL of the type ‘number approximation’ is given below (Table 1).
Table 1: Most frequently occurring number approximations, in descending order

<table>
<thead>
<tr>
<th>VL item (Overall number of occurrences in brackets)</th>
<th>NSE</th>
<th>NSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>about n (225)</td>
<td>about n (128)</td>
<td>about n (97)</td>
</tr>
<tr>
<td>n or m* (167)</td>
<td>n or m* (102)</td>
<td>n or m (65)</td>
</tr>
<tr>
<td>maybe n (70)</td>
<td>maybe (37)</td>
<td>n to m (33)</td>
</tr>
<tr>
<td>n to m (35)</td>
<td>n to m (17)</td>
<td>maybe (18)</td>
</tr>
<tr>
<td>Overall %</td>
<td>55%</td>
<td>45%</td>
</tr>
</tbody>
</table>

* Note: n = first number, m = second number e.g. 15 or 20

The rank order here is identical except for maybe and n to m, whose rankings are reversed. The rankings for all other exemplars were very similar, which suggests that there are very real similarities between the VL usage of NSE and NSC participants, especially in the matter of preference for certain VL exemplars. While on the whole the NSE participants use more VL, there is no indication that they are using items which the NSC do not know. One explanation for this is that Cantonese has a large number of vague quantifying and other expression expressions: in a recent focus group discussion on this issue, I found that all but the most obscure and infrequent English VL has a Cantonese equivalent. The one interesting difference is that the NSC use ‘etcetera’ as a vague category marker while NSE do not. One might speculate that NSC carry over this genre-inappropriate item (a written, rather academic lexical item) because of the greater emphasis placed on writing in the Hong Kong education system (cf. De Cock et al. 1998).

5.3 Patterns: collocation and clustering

To determine whether it would be a useful approach for all the data, I decided to look at the collocational patterning associated with the most frequently used VL items by selecting the top two numerical and non-numerical approximators, as well as the most common category marker and placeholder. It was thought that this might illuminate interesting similarities or differences in the kinds of things which are ‘vaguified’ (collocations) and whether the two sets of speakers used similar vague phrasing (clusters). This exercise might be described as a piloting stage designed to determine whether it might be useful to adopt this approach for all VL exemplars. The findings are reported below, beginning with the numerical approximator about.
5.3.1 About

It was found that the type of lexical item which most often immediately follows about is a simple number (in the range 1–30) for both groups of participants, with the numbers 2 and 1 being the most common for NSE and NSC respectively. In the second position to the right of about the most common items are usually either time lexis (‘years’, ‘minutes’, ‘hours’) or a larger, base-ten enumerator (‘100’, ‘1000’). The concordancing program also revealed the relatively high frequency of the phrase about the same for both NSE and NSC (Table 2).

Table 2: Clusters about

<table>
<thead>
<tr>
<th>NSE</th>
<th>NSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>about the same</td>
<td>about the same</td>
</tr>
<tr>
<td>about an hour</td>
<td>for about one</td>
</tr>
<tr>
<td>about two hours</td>
<td>about I think</td>
</tr>
<tr>
<td>about two years</td>
<td>about ten years</td>
</tr>
</tbody>
</table>

Both NSE and NSC display some use of idiomatic patterns (clusters), the most frequent being the same for both sets of speakers. All the NSE clusters, apart from the first, concern periods of time, whereas the NSC clusters are more varied.

5.3.2 N or m

The second most frequent numerical approximator was n or m (where n and m represent numbers). The collocation patterns are very similar for both sets of speakers and concern periods of time and money, with references to dollars occurring in both the first and second positions to the right of the vague expression. For both groups of speakers, right collocates which immediately follow n or m either refer to periods of time (‘years’, ‘days’, ‘months’) or round, base-ten enumerators (‘100’, ‘1000’), supporting Channell’s (1990: 101) claim that reference point (round, particularly base ten) numbers are more likely to be vaguified than others. In the second position to the right of n or m, ‘ago’ commonly occurs, indicating that this expression is often used to manage recall of events in the past.

Channell (1994: 56) identifies constraints on the kinds of numbers which may be combined using n or m, suggesting, for example, that where the n number is a multiple of 10, the m number will be 10 above. She also notes that the difference between the two numbers increases along with the size of the numbers. The clusters here (Table 3) seem to conform to the suggested norms apart from ten or eleven, but it is easy to imagine a circumstance in which this would be perfectly acceptable.
Table 3: Clusters *n or m*

<table>
<thead>
<tr>
<th>NSE</th>
<th>NSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>two or three</td>
<td>three or four</td>
</tr>
<tr>
<td>three or four</td>
<td>five or ten</td>
</tr>
<tr>
<td>four or five</td>
<td>one or two</td>
</tr>
<tr>
<td>an hour or</td>
<td>or thirty dollars</td>
</tr>
<tr>
<td>one or two</td>
<td>ten or eleven</td>
</tr>
<tr>
<td>six or seven</td>
<td>ten or twenty</td>
</tr>
<tr>
<td>half an hour</td>
<td>two or three</td>
</tr>
</tbody>
</table>

5.3.3 *A lot of*

This was the most frequent non-numerical approximator. The most frequent collocations were ‘people’, ‘money’, ‘things’ and ‘work’ for both groups, with the addition of ‘places’ for the NSC. Clusters were also very similar (Table 4). Clearly, there are grammatical and other similarities here, such as the head nouns relating to people and money, the only difference being the NSC reduplication in *lots and lots of* and the higher NSE ranking for the (hedged) *quite a lot of*.

Table 4: Clusters *a lot of*

<table>
<thead>
<tr>
<th>NSE</th>
<th>NSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>a lot of people</td>
<td>a lot of people</td>
</tr>
<tr>
<td>quite a lot of</td>
<td>a lot of er</td>
</tr>
<tr>
<td>have a lot of</td>
<td>a lot of money</td>
</tr>
<tr>
<td>a lot of money</td>
<td>have a lot of</td>
</tr>
<tr>
<td>and a lot of</td>
<td>lots and lots of</td>
</tr>
<tr>
<td>a lot of the</td>
<td>quite a lot of</td>
</tr>
</tbody>
</table>

5.3.4 *Many*

This was the second most common non-numerical approximator. ‘Not’ as a left collocate ranks more highly for the NSE than for the NSC group. It is perhaps overly formal to use *many* as a determiner in a positive statement, except where premodified (Berry 1997: 106). The concordancer revealed that it is far more likely to be used in the negative (‘not many’), in questions (‘how’) or with the intensifier ‘too’: there were no other collocates for the NSE, but in the NSC data, the verb ‘have’ also appeared as a collocate. One possible conclusion to be drawn from this is that the NSC may be less aware of usage constraints. However, the fact that the phrase ‘I could have many’ occurs as a common NSE cluster shows that this conclusion may be unwarranted (Table 5). Right collocates were identical for both groups: ‘people’, ‘things’, ‘years’, ‘times’, ‘kinds’ and ‘choices’ (in rank order).
Table 5: Clusters many

<table>
<thead>
<tr>
<th>NSE</th>
<th>NSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>there are so many</td>
<td>I could have many</td>
</tr>
<tr>
<td>there are many many</td>
<td>I don’t know how</td>
</tr>
<tr>
<td>I could have many many</td>
<td>many many kinds of</td>
</tr>
<tr>
<td>I don’t know how many</td>
<td>many many times I have</td>
</tr>
<tr>
<td>many many kinds of</td>
<td>many times I have</td>
</tr>
</tbody>
</table>

5.3.5 Partial specifiers, category markers and placeholders

There were no significant correlations for partial specifiers (more, less), vague category markers or placeholders. This may be due to the small number of exemplars of each type but perhaps also demonstrates that these forms can be used with all topics; they are freely combinable and give rise to fewer fixed patterns. The fact that they are relatively semantically empty makes them less amenable to concordancing analysis than more specific lexemes. Table 6 presents the clusters for the most common category marker, or something.

Table 6: Clusters or something

<table>
<thead>
<tr>
<th>NSE</th>
<th>NSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>or something like that</td>
<td>or something like that</td>
</tr>
<tr>
<td>or something of that</td>
<td></td>
</tr>
<tr>
<td>something of that nature</td>
<td></td>
</tr>
</tbody>
</table>

There is some difference in the clustering here with the NSE using a cluster which the NSC do not. Table 7 presents clusters for the most common placeholder, thing.

Table 7: Clusters thing

<table>
<thead>
<tr>
<th>NSE</th>
<th>NSC</th>
</tr>
</thead>
<tbody>
<tr>
<td>that kind of thing</td>
<td>that kind of thing</td>
</tr>
<tr>
<td></td>
<td>that sort of thing</td>
</tr>
<tr>
<td></td>
<td>this kind of thing</td>
</tr>
</tbody>
</table>

The NSC display more variety in this case. No clusters occurred more than twice for sort of.

Once again, there would seem to be a remarkable similarity in the most frequent collocates, suggesting that, at least for these VL items, NSE and NSC are being vague about similar things. Number approximations naturally (by definition) expect a following number, but in the R2 position (i.e. two words to the right of the node word) one would expect considerable variation since it can be occupied by
anything which can be quantified. Yet the most frequently occurring items are almost identical for the two sets of speakers; they overwhelmingly concern time, people, places, work and money. There does not seem to be any support for the claim that fixed expressions are more fixed for native than they are for non-native speakers (contra De Cock et al. 1998).

On this evidence, quantification using concordancing software is useful for getting an overall picture of how much VL is used by each type of speaker (NSE, NSC). However, an emphasis on frequency, as is necessitated by a quantitative approach to data analysis, can mask potentially interesting infrequent cases, and overemphasise repeated combinations of words in the immediate linguistic context. Since concordancing is based on high frequencies, it tends to focus on popular (i.e. high frequency) items, which might give learners of a language fewer problems because there will have been more exposure to them during the language learning process. Problems in VL usage might therefore be hidden by a concentration on high frequency items. Again, given that several types of VL are semantically empty and grammatically unrestricted, much of importance is also not amenable to collocational analysis. To find out more about how NSE and NSC actually use VL, it is necessary to look in more detail at the linguistic context of utterances, but with the attitude that even single instances can be revealing about speaker attitudes and motivations. Lack of space precludes a detailed discussion of all the VL in the corpus, but as an illustration of this more qualitative approach I use the example of stuff in the following section.

6. Usage: the example of stuff

Although the figures presented above seem to suggest a great deal of similarity in usage, we need to look more deeply into the functions and usage of VL before drawing conclusions. To that end, I examine below one VL exemplar, the category marker and placeholder stuff. The range of functions, epistemic and affective, performed by the VL in the NSE and NSC data are presented in order to illustrate the differences in usage between the two groups of speakers. There is obviously a limit to what can be understood from a single line of concordance printout, but I hope that this will be enough for the reader to at least get a sense of what is happening. The NSE participant is identified by use of a capital letter and the NSC by the use of a lower case letter.

6.1 NSE (54 occurrences)

The NSE uses stuff in a wide range of linguistic environments (with nouns, in lists, attached to clauses and so on), and as a placeholder

(1) which is quite nice if you like juicy stuff a: right B: but yeah I’m gonna h

and a category marker:

(2) B: some people got aunts and uncles and stuff there a: no no no I know I don’t k
There is ample evidence of *stuff*’s usefulness as a compensator for linguistic inadequacy (or lack of knowledge) for example:

(3) this a a from six bottom you mix tiger *stuff* shark and snake in it is quite exp

The speaker may not have a clear idea of what he is talking about, or does not have the appropriate vocabulary, since the topic is Chinese medicine.

There is reference to categories that are readily accessible (though not necessarily lexicalised in English), for example clothing:

(4) walking around in shorts and T-shirts and *stuff* b: yea they are local people but

The mention of two category members allows one to infer which category is meant, perhaps ‘very casual wear’.

*Stuff* compensates when the speaker lacks specialised scientific terminology:

(5) now you know because there’s this beef *stuff* that you can get [diseases from [t Encephalitis, BSE and ‘Mad Cow Disease’ are all possibilities here: the speaker has either forgotten or does not know the correct term.

VL can be useful as a circumlocutory device by a speaker who is not quite sure of what she wants to say:

(6) ul of sugar a: ((laughs)) B: all that *stuff* er you know when you look at for t

Several speakers use *stuff* to create what might be called fuzzy, ad hoc categories (after Barsalou, cited by Channell, 1994: 122):

(7) n something with their B: Bible memory *stuff* and see how we could  a: really

Notice that even though there is no obvious category invoked by ‘Bible memory *stuff*’, there is no communication breakdown and speaker ‘a’ seems to understand what is meant.

The adjunctive *stuff* is used to convey the sense that there is more that might be said:

(8) about women when they’re together and *stuff*= B: =and I know men are like

Speaker B is making some general comments on male-female relationships and feels that she does not have to go into greater detail to make her point.

Another use found here is for de-emphasising an item of information for rhetorical purposes so as to emphasise contrasting (more precise or well explained) information:

(9) the lunch they will simply to have some *stuff* fill their stomach. For dinner you
The contrast here is between lunch, a quick meal where getting full is the priority, and dinner, more leisurely and enjoyable.

*Stuff* can be used to convey a variety of stances such as incredulity,

(10) r the world or where do they get this **stuff** this **stuff** b: I don’t know from US

as well as defensiveness,

(11) Nike (.) I’m not very good on that **stuff** A: anyway he went there after school

for deference or politeness through mitigation or softening, as in this case where potentially face-threatening advice is given,

(12) g right A: there is a lot of things or **stuff** you need to learn about B: quite

And pejoratively, often for humorous dismissiveness:

(13) A: I mean I don’t wear that sportswear **stuff** particularly c: but there’s a time

*Stuff* occurs at certain points within the discourse where the speaker takes stock of the general trend of what has been said and attempts to generalise. This is often preceded by a deictic:

(14) class so (.) mm A: you need to do that **stuff** I think= b: =mm (.) huh A: but ar

*Stuff* is also used in combination with deictics when the intended referent is easily identifiable:

(15) the case  B: yea I would it leave that **stuff** on the floor  d: you you won’t get

Finally, there are many examples of what Hyland (1998: 150) calls ‘harmonic combinations’, where VL co-occurs with hedging or other VL types, for example ‘kind of’:

(16) t you you took A yea they keep track of **stuff** [ so it’s kind of a neat system b

Thus there is tremendous variety in the uses to which *stuff* is put. It is grammatically flexible and can be used to convey, or reinforce, many different emotions. Overstreet (1999: Ch. 5) demonstrates that VL is useful not only epistemically for categorising and giving the ‘right’ amount of information but also for a range of interpersonal functions which suggest that speakers share knowledge and assumptions. The message is that specificity is unnecessary, and this suggestion of mutual familiarity enhances conversational rapport. The ‘basic’ use may be to compensate for lexical or other deficiencies (placeholder function) or to suggest that one could say more but prefers to suggest it instead, but *stuff* is obviously multi-dimensional. It does not work in
isolation of course, but achieves these effects in combination with other verbal and non-verbal messages.

6.2 NSC (19 occurrences)

There are not as many examples from the NSC data so one would not expect such a great variety of functions. However, even taking this into account, the relative homogeneity of the data is noteworthy, as is the fact that *stuff* is never used as a category marker.

There is often a lexical item or phrase to replace *stuff*, for example:

(17) hose and [anatomical [and physiological *stuff*] A: [yeah] [yeah

*Stuff* could easily be replaced with ‘subjects’ here and does not seem to convey any sense of dismissiveness.

Sometimes an acceptable word (synonym) occurs within the lexical environment:

(18) a: =I I I like analysing data analysing *stuff* I like this so I _ I I just try my

In some cases there is linguistic (agreement) error,

(19) use) b: shall I go and get one of those *stuff* erm anyway A: yea okay thanks (pau or redundancy (or at least nothing added to the core meaning):

(20) ax it off anyway m: how about the cover *stuff* do you have any A: yes [that’s

The speaker is referring to the cover of a magazine so it is difficult to see what the addition of *stuff* adds to the message here. Once again, there is no reason to believe that the speaker is using the word pejoratively.

The use of *stuff* in the following example could be compensation for lack of intimate knowledge of the nature of the business (self-protection, since her elaborations are equally vague):

(21) firm yeah (.) working on communications *stuff* (.) so ah B: like what a: like t

On this evidence, it seems that NSC do not employ the full range of functions for which *stuff* can be used. Scrutiny of these utterances in context reveals that they are very much epistemic uses of VL and seem relatively unmotivated by affective concerns.

Similar findings have emerged in a preliminary look at occurrences of *sort of*. All of the functions of *sort of* noted by Aijmer (1984) occur in the NSE data, but only two can be found in the NSC examples. Holmes (1988) identifies epistemic and affective meanings of *sort of*, the latter within the tradition of politeness research, all of which are in evidence in the speech of the NSE. In contrast, the NSC make less
affective use of *sort of* and there seem to be fewer harmonic combinations. Holmes’s (1988) conclusion is that discourse type rather than speaker sex (her original object of investigation) largely determines the frequency of occurrence. My work suggests that, for intercultural interaction, another variable may be the native language of the speaker. The nature of these influences remains to be explored in greater depth.

7. Conclusion

On the strength of this preliminary work, it seems that the figures on preferred types of VL suggest a greater similarity than do the overall quantities (which show clearly that NSE participants are ‘vaguer’ than NSC), and the more detailed functional analysis. It is in this last area where the NSE emerges as the more skilled user of the language, making full use of this flexible linguistic resource. Note that there is little evidence here of NSE accommodation or of NSC communication strategies. There is also no evidence that the NSC inaccuracies and limitations hinder conversational cooperativeness or prevent the achievement of conversational goals, and while there are linguistic problems it is rather difficult to say whether there are also pragmatic or sociopragmatic ones (Thomas, 1983: 91).

Various explanations are possible for the differences noted above. One is that the first language of the NSC participants interferes with their performance in English. This seems unlikely since Cantonese has a range of expressions which are functionally equivalent to both the placeholding and category marking functions of *stuff* (D. Li, personal communication 14 February 2001). There may of course be register restrictions on such forms in the first language which might preclude their use in the situations which arise in these data. Another possible explanation is an educational one. Perhaps it is the case that prevailing pedagogical methods do not allow for sufficient exposure to native language models which contain VL, such as informal conversation. Proving that these factors influence the acquisition of VL is a complex undertaking however, requiring the collection and analysis of additional data, and so is outside the scope of the present study.

For the next phase of the research it will be necessary to complement the approaches and methods described above, where the NSE and NSC data are essentially being treated as separate corpora, with one more in line with current thinking about best practice in the analysis of conversation. There is a need to take into account participant reactions to utterances (Schegloff 1992: 101) and to look at conversations in their entirety. The question is not only which *linguistic* environments cause people to use vague language, but which other environments do so, how these environments become instantiated and with what consequences. An interpretative framework is needed to aid understanding of a lexical item, and this may be very difficult to assemble from concordance lines alone: each line may have to be expanded to take into account larger chunks of discourse, up to and including the whole conversation. There can be no doubt that corpora of naturally occurring data are the best way forward for studies of conversation and, as de Beaugrande (1996) argues, it would be difficult to see how meaningful pragmatics research, which focuses on the various functions performed by utterances in different contexts, could be conducted in any other way.
Notes

1 The research is being undertaken for a Ph.D. at City University of Hong Kong, under the supervision of Professor Chris Candlin.

2 For simplicity, the abbreviations ‘NSE’ and ‘NSC’ refer to both single speakers and groups of speakers.

References


