

Translating Vagueness? A Study on Translations of Vague Quantifiers in an English-Chinese Parallel Corpus

Amy Y. Wang¹ and Scott Piao²

Abstract

In this paper we investigate variation in the translation of three vague quantifiers, *many*, *some* and *a few* between English and Chinese. Studies of ‘linguistic vagueness’ (sometimes called ‘language vagueness’ in previous research) regard vagueness as a general phenomenon in language. In this type of study, vagueness is often discussed – but not limited – to *indeterminacy* of the referential boundary of words. Previous studies of vagueness have generally followed two related – but different – approaches (i) vagueness viewed as pervasive in language and (ii) vagueness viewed as referring to certain expressions that are considered as vague. We refer to the first type as the study of linguistic vagueness, and the second type as the study of pragmalinguistic vagueness. In this study, we focus on the second perspective of vagueness. We use an English/Chinese parallel corpus that contains newspaper articles and literary reviews to explore the translation strategies. Our study shows that each quantifier has a typical translation that follows the scalar implicature in both original and targeted languages. The atypical Chinese variants may be used to translated the same English quantifier; that means, the corresponding translation variant may not necessarily reflect the quantity conveyed in the quantifier(s) of the original language (i.e. English).

1. Introduction

Vagueness is a long-standing research topic that, in modern times, can be traced back to Peirce (1901) and Russell (1923). Previous studies of vagueness (Williamson, 1994a; Channell, 1994; Jucker et al., 2003) have generally followed two related – but different – approaches: (i) vagueness viewed as pervasive in language, and (ii) vagueness viewed as referring to certain expressions that are considered as vague. We refer to the first type as the study of linguistic vagueness, and the second type as the study of pragmalinguistic vagueness. However, in so far as some studies of vagueness make use of both concepts it is useful to see them as being at the ends of a continuum of vagueness in language, separated by the extent to which they differ in terms of (i) the degree of pervasiveness of vagueness in language and (ii) the words that they define as instances of vagueness. See Figure 1 below.

¹ Lancaster University/Manchester Metropolitan University
e-mail: y.wang@mmu.ac.uk

² Text Mining Group, School of Computer Science, The University of Manchester
e-mail: scott.piao@manchester.ac.uk

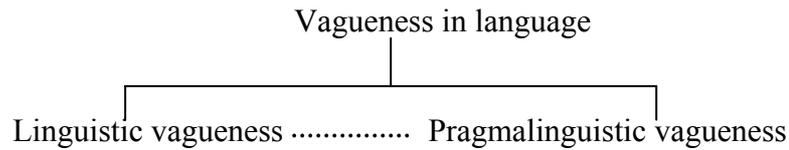


Figure 1: The two main types of studies of vagueness in language

Studies of ‘linguistic vagueness’ (sometimes called ‘language vagueness’ in previous research) regard vagueness as a general phenomenon in language. In this type of study, vagueness is often discussed in terms of – but not limited to – *indeterminacy* of the referential boundary of words. This concept was first commented on by Eubulides of Miletus (in Ancient Greece), and has become known as ‘sortes paradox’. Famous examples of sortes paradox are the words *heap* and *bald*. How many grains do we need in order to have a *heap*? How bald is *bald*? Consider the following utterance, which is often discussed in the literature on vagueness in language:

Example 1: John is bald.

Some people regard the utterance as vague because it is not clear how few hairs John is expected to have to be considered bald (e.g. Eubulides, cited in Williamson, 1994a; Russell, 1923; Williamson, 1994a). Is it less than 2, or 20, or 200, or 2000 hairs? A related question to this is the distribution of John’s hairs (cf. Williamson, 1994a).

In contrast, studies of ‘pragmalinguistic vagueness’ (known as ‘vague language’ in previous literature) only focus on specific expressions that are themselves vague and/or which add vagueness to utterances (e.g. Channell, 1994; Drave, 2002). Examples include vague quantifiers (e.g. *a lot of*, *many*), vague approximators (e.g. *about*), and placeholder words (e.g. *thingy*, *thingummy*).

A major difference between the study of linguistic vagueness and the study of pragmalinguistic vagueness is the choice of words or phrases taken to exemplify vagueness. Examples that are commonly discussed in the former type of study are nouns (e.g. *heap*) and adjectives (e.g. *tall*, *hot*, *bald*). However, in pragmalinguistic studies, the examples discussed are usually adverbs, adverbial elements or determiners (e.g. *many*, *a lot*, and *things like that*). Our analysis of vague quantifiers belongs to the latter type of study.

Vague quantifiers (e.g. *many*, *several*, *a few*) in previous research have been discussed with respect to at least three different approaches. The first approach is a formal semantic approach. It treats the quantifications conveyed in vague quantifiers in terms of “the mapping of elements or sets onto other sets” (Moxey and Sanford, 1997: 207). In some studies of this approach, natural language quantifiers are viewed as “a subset of theoretically possible quantifiers, and part of the interest lies in determining what are the restrictions on natural language quantifiers with respect to the possibilities generated by a particular theoretical framework” (Barwise and Cooper, 1983; Westersthål, 1989 cited in Moxey and Sanford, 1997: 207). The second approach discusses vague quantifiers on the basis of the hypothesis that certain proportions or numbers can be assigned in explaining the meanings of these quantifiers (e.g. Mosteller and Youtz, 1990; Clark, 1990; Channell, 1994). The discussions concerning vague quantifiers in this approach are often centred on the scaling of the quantities conveyed in these expressions (see also discussions in Moxey

and Sanford 1997). The third approach is a functional one, which aims to trace the reasons for using using different quantifiers in different communicative situations. Moxey and Sanford, who are the main representatives of this approach, suggest that vague quantifiers “do not simply denote amounts” (Moxey and Sanford, 1997:208), and they further suggest that the choice of quantifiers is related to speakers’ subjective assumptions.

Our approach differs from most previous studies in that we investigate the vagueness issue using a corpus linguistic methodology. In this paper we investigate variation in the translation of three vague quantifiers, *many*, *some* and *a few*, between English and Chinese, using corpus data. We used two English-Chinese parallel corpora that contain newspaper articles (approx. 595,000 tokens) and literary reviews (approx. 107,000 tokens) respectively³.

The remainder of this paper is organised as follows. Firstly we discuss the issue of translating vagueness between English and Chinese. We then carry out an investigation of the following research questions:

- 1) Is it possible in all cases to translate vague quantifiers?
- 2) What are the possible variants in the translation?
- 3) Does genre have an effect on the choice of translation variant?
- 4) What other factors influence the choice of translation variant?

2. Translations of vague quantifiers

In translating vague quantifiers from one language to another, between Chinese and English in our particular case, it may be difficult to precisely map the correspondences of scalar implicature in the two languages. As shown in Examples 2 and 3, there are two sets of quantifiers in Chinese and English that can be translations of each other in certain contexts.

The scalar implicature that we refer to here is the relations between words and their properties (e.g. quantity, degree of certainty, degree of possibility) within individual languages. For instance, in most contexts, the majority of native English speakers would interpret the quantity denoted by *very many* to be greater than that of *many*; the quantity conveyed by *many* greater than that of *some*, and the quantity of *some* greater than that of *a few*, and so on. The same phenomenon is also apparent for native Chinese language speakers. However, when translating such context-dependent quantifier words between these two languages, it is difficult to pinpoint precisely where they correspond to each other. In Examples 2 and 3, does *hěnduō* (literal translation: very many) denote the same quantity as *very many* or *many* in English? Does *yìdiǎndiǎn* correspond to *a few* or *few* in English?

³ The newspaper corpus is provided by the CCID (China’s Center of information Industry Development) and the literature commentary articles is provided National Research Centre for Foreign language Education (NRCFLE) of Beijing Foreign Studies University, China..

English:	very many	many	some	a few	few
Chinese :	<i>hěnduō</i> ⁴	<i>duō</i>	<i>yìxiē</i>	<i>yìdiǎn</i>	<i>yìdiǎndiǎn</i>
	很多	多	一些	一点	一点点

Example 2: Possible mapping of vague quantifiers in English and Chinese based on the default lexical meaning

English :	very many	many	some	a few	few
Chinese :	<i>hěnduō</i>	<i>duō</i>	<i>yìxiē</i>	<i>yìdiǎn</i>	<i>yìdiǎndiǎn</i>
	很多	多	一些	一点	一点点

Example 3: Possible mapping of vague quantifiers in English and Chinese based on the default lexical meaning

3. Case study: frequency of the translation variants of vague quantifiers in the corpus

In order to investigate the issue of how English vague quantifiers can be translated in Chinese, we analysed the frequencies of the Chinese translation variants for each of the three English expressions: *many*, *some* and *a few*. Performing the frequency analysis involved the following steps: 1) the Chinese/English parallel corpus data was aligned at sentence level; 2) sentence pairs which contain any of the three English vague quantifiers were extracted; 3) the sentence pairs were manually checked through to mark-up the correct Chinese translations; 4) frequency information for the translations of each English vague quantifier was collected.

3.1 Frequency of ‘many’

Figures 3 and 4 show the frequencies of the main translation variants of the English vague quantifier, *many*. The first line lists common Chinese variants of this English quantifier that occur in the corpus data. Here we constrain our investigation to the translation given by the translator of the text, rather than our own translation, in order to minimize the use of subjective judgement in identifying the translations. The Chinese quantifier translations are sorted in ascending order from left to right by the size of quantity conveyed by each word. For example, *xǔduō* on the left end indicates the smallest amount in the list while *zhòngduō* on the right end indicates the greatest amount. The figures in brackets indicate the frequencies of the translations. In

⁴ The Pinyin is used to transcribe the sounds of the Chinese words, and the tone symbols are marked on top of the vowels.

addition, each Chinese word is given its spelling in Pinyin, i.e. the phonemic spelling system of the (Mandarin) Chinese language, which is used to represent the pronunciation of the words.

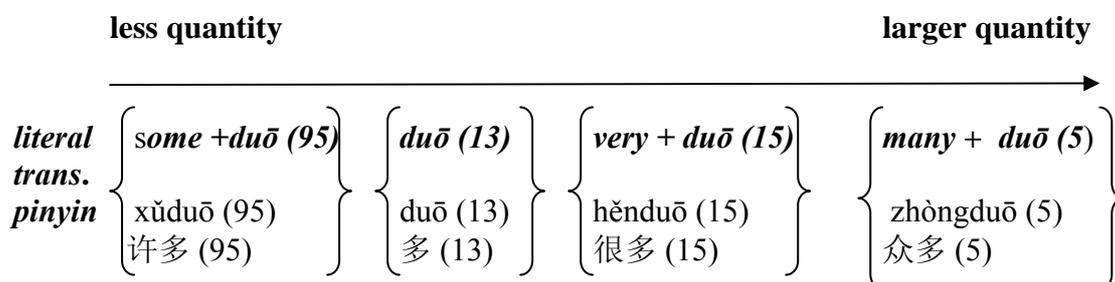


Figure 3: Newspaper

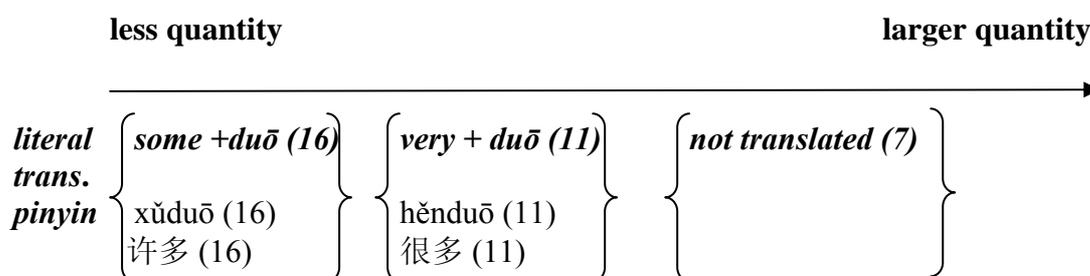


Figure 4: Literary commentary texts

As shown in Figures 3 and 4, the Chinese translation variations of *many* include *xǔduō* (some + *duō*), *duō*, *hěnduō* (very + *duō*) and *zhòngduō* (many + *duō*), of which the morpheme ‘*duō*’ (corresponding to ‘many’), is the basic element of each translation. Combined with modifier morphemes, it indicates various quantities. In most contexts, the quantity of *hěnduō* is interpreted by native Chinese speakers to be more than that of *duō*, and *duō* more than *xǔduō*, and so on.

Regarding the frequencies of the Chinese translations, *xǔduō* (some + *duō*) appears to be the most common translation in both of the genres of newspapers and literary commentaries, evident from 95 occurrences in the former and 16 occurrences in the latter. The second most frequent translation is *hěnduō* (literal translation *very + duō*), with 15 and 11 occurrences in newspapers and commentaries respectively.

It can be noted that, in some contexts, the quantifier *many* is not translated, and this is especially the case in the literary reviews (7 occurrences), as illustrated in Example 4 below:

Example 4:

Her first novel was to be *The Professor* based on **many** of her experiences, as was her later *Villette* and even Shirley sees her create a hero in Robert Gerard Moore who is half Belgian

她	寫的	第一部	小說	《教授》	就基於	
Her	write	first	novel	Professor	based on	
{NULL}	自己的	親身經歷，	稍後的	《維萊特》	也是，	就連
	onself's	own experience	later	Villete	also be-V	even
《謝利》	裏面	也	有	埃熱的	影子：主人公	
Shirley	in(side)	also	have		shadow main character	
羅伯特·吉羅德·爾	是	半個	比利時人。			
Robert Gerard Moore	is	half	Belgian			

她寫的第一部小說《教授》就基於 {NULL} 自己的親身經歷，稍後的《維萊特》也是，就連《謝利》裏面也有埃熱的影子：主人公羅伯特·吉羅德·爾是半個比利時人。

There are also some other cases where the meaning of the quantifier is incorporated into the meaning of the noun that follows. For instance, ‘many kinds’ is translated as 丰富 fēngfù (literary means *rich* in English). ‘Many kinds’ can also be translated as 种种 zhǒngzhǒng, literary meaning *every kind*.

3.2 The case of ‘some’

The vague quantifier ‘some’ has different translation variants in Chinese, as the tables below indicate. Table 1 presents the frequencies of the variants in the newspaper texts, and Table 2, the literary reviews.

Chinese Pinyin	English (literal)	Freq.	许多	xǔduō	<i>many</i>	2	
一些	yìxiē	<i>some</i>	107	一项	yíxiàng	<i>one</i>	1
某些	muǒxiē	<i>certain+some</i>	26	一定的	yídingde	<i>certain</i>	1
约	yuē	<i>about</i>	10	一点	yìdiǎn	<i>a bit</i>	1
某种	muǒzhǒng	<i>certain+type</i>	6	一直	yìzhí	<i>constantly</i>	1
部分	bùfèn	<i>part</i>	8	些许	xiēxǔ	<i>some +little</i>	1
一段	yíduàn	<i>a part</i>	4	其它	qítā	<i>other</i>	1
一些人	yìxiērén	<i>some people</i>	4	左右	zuǒyòu	<i>around</i>	1
有人	yǒurén	<i>certain person</i>	4	某项	muǒxiàng	<i>certain</i>	1
某个	muǒge	<i>certain one</i>	4	某	muǒ	<i>certain</i>	1
大约	dàyuē	<i>about</i>	3	某些人	muǒrén	<i>certain person</i>	1
有些	yǒuxiē	<i>there is some</i>	3	点	diǎn	<i>a bit</i>	1
这些	zhèxiē	<i>this+ some</i>	3				
一部分	yíbùfèn	<i>a part</i>	3				

Table 1: Frequencies of the translation variants of some in the newspaper texts

Chinese	Pinyin	English literal		任何	rèhé	<i>Any</i>	1
NULL	not translated		15	其它那些	qítānàxiē	<i>other+those</i>	1
一些	yìxiē	<i>some</i>	7	几首	jǐ shǒu	<i>several+classifier</i>	1
有些	yǒuxiē	<i>there is some</i>	3	大约	dàyuē	<i>About</i>	1
某种	muǒzhǒng	<i>certain type</i>	6	好几	hǎojǐ	<i>very+several</i>	1
一部分	yíbùfèn	<i>a part</i>	2	无数	wúshù	<i>Many</i>	1
几个	jǐ ge	<i>one+classifier</i>	2	有	yǒu	<i>there is</i>	1
某些	muǒxiē	<i>certain+some</i>	2	有些人	yǒuxiērén	<i>some people</i>	1
一棵	yì ke	<i>one+classifier</i>	1	有些故	yǒuxiē	<i>some stories</i>	1
一种	yì	<i>one+classifier</i>	1	事	gùshì		1
一个	yí ge	<i>one+classifier</i>	1				
一些酒	yìxiējiǔ	<i>some+wine</i>	1				

Table 2: Frequencies of the translation variants of *some* in the literary reviews

The typical translation of ‘some’ in the newspaper texts is ‘yìxiē’, with 107 occurrences. However, in the literary reviews, ‘some’ tends not to be translated (15 occurrences), and ‘yìxiē’ is used only 7 times. This suggests that the choice of translation variant in the English-Chinese texts may be sensitive to genre. Second, in addition to its most typical translation in Chinese, ‘some’ can be translated as *xǔduō* (2 occurrences). *Xǔduō* is the major corresponding variant of the quantifier ‘many’, as discussed in section 3.1. This suggests that in translating the English quantifiers into Chinese, the Chinese corresponding quantifiers do not necessarily reflect the quantity conveyed in the original quantifier: two quantifiers that imply different quantities (e.g. *some*, *a few*) can be translated to the same corresponding form (e.g. *yìxiē*).

3.3 The case of ‘a few’

The following two tables present the frequencies of the translation variants of *a few*, in the newspaper and literary reviews.

Chinese	Pinyin	English (literal)	Freq.
几	jǐ	several	7
一小撮	yìxiǎocuò	a small amount	4
一小部分	yìxiǎo bùfèn	a small part	2
一些	yìxiē	some	2
数	shù	several	2
几枚	shùméi	several +classifier	1
数日	shùrì	several +classifier	1

Table 3: Frequencies of the translation variants of ‘a few’ in the newspaper texts

Chinese	Pinyin	English	Freq.
几	jǐ	several	3
一些	yìxiē	some	2
NULL	not translated		1
几个	jǐ +classifier	several	1
几根	jǐ +classifier	several	1

Table 4: Frequencies of the translation variants of ‘a few’ in the literary reviews
 First, as shown in Table 3 and 4, *a few* is typically translated as *jǐ*, which is sometimes used as a corresponding form of *several* in English. Second, the second most frequent translation variant is different between the two genres. *Yìxiǎocuò* occurs four times in the newspaper texts, while *yìxiē*, sometimes used as a corresponding form of *some*, is used twice in the literary reviews. This suggests a similar tendency to that of ‘many’ and ‘some’, in that the typical corresponding variant is the same in both genres, and the atypical translation may be used to translate two different English quantifiers. For instance,

4. Further discussion and implication

As the data shows, while there is no fixed translation variant for any of the three quantifiers, *many*, *some* and *a few*, in English and Chinese, nevertheless each quantifier has a typical corresponding variant (i.e. the most frequent form), and a set of possible atypical (i.e. less frequent) corresponding variants. For example, in the case of *many*, the typical corresponding variant is *xǐduō*, and the other less frequent variants include *duo*, *hěnduō*, *zhòngduō*. As to the typical translation of the three

quantifiers, it is noticeable that they follow the scalar implicature of the corresponding form in the source language. However, with regard to atypical translation variants, they, in some cases, break the scalar implicature of the original language (i.e., English) because two English quantifiers may be translated into the same Chinese quantifier. This seems to suggest that in such cases, the quantifiers in the targeted language do not necessarily reflect the quantity conveyed in the original quantifier. A further finding to note is that the quantifiers are not always consistently translated. This is particular the case in the literary reviews, where ‘some’ is frequently not translated at all.

4. Conclusion

Our analysis of variation in translations of three quantifiers, *many*, *some* and *a few*, in the parallel corpus of English and Chinese found that each quantifier has a typical corresponding form in the target language (i.e. Chinese). These typical forms maintain the scalar implicature that is present in the source language (i.e. English). While the atypical translations may overlap between the quantifiers, they do not represent the absolute quantity of the original quantifiers. Further research is required to investigate the factors that influence the choice of the corresponding variant in the target language, as well as the factors that affect the untranslatability of vague quantifiers between languages.

References

- Channell, J. 1985. Vagueness as a conversational strategy. *Nottingham Linguistic Circular* 14: 3–24.
- Channell, J. 1990. Precise and vague quantities in writing on economics. In *The writing scholar: studies in academic discourse*, ed. W. Nash, pp. 95–117. London: Sage.
- Channell, J. 1994. *Vague language*. Oxford: Oxford University Press.
- Jucker, A.H., S.W. Smith and T. Lüdge. 2003. Interactive aspects of vagueness in conversation. *Journal of Pragmatics* 35: 1737–69.
- Moxey, L.M. and A. Sanford. 1997. Choosing the right quantifier: usage in the context of communication. In *Conversation: Cognitive, communicative, and social perspectives.*, ed. Givón T., pp. 207–231. Amsterdam: John Benjamins.
- Peirce, C.S. 1902. Issues of pragmaticism. In *Collected writings of Charles Sanders Peirce*, ed. P. Hartshorne, P. Weiss and A. Burk. Cambridge: Harvard University Press.
- Russell, B.A.W. 1923. Vagueness. *Australian Journal of Philosophy and Psychology* 1: 84–92.
- Wang, A.Y. 2007. Vagueness in interaction: a cross-cultural examination of vagueness in Taiwanese and British courtroom discourse. PhD thesis, Lancaster University.
- Williamson, T. 1994a. *Vagueness*. London: Routledge.
- Williamson, T. 1994b. Vagueness. In *The Encyclopedia of Language and Linguistics*, ed. R. Asher, pp. 4869–71. Oxford: Pergamon.