Understanding the Use of Quantifiers in Mandarin

Guanyi Chen and **Kees van Deemter**
Department of Information and Computing Sciences
Utrecht University
{g.chen, c.j.vandeemter}@uu.nl

Abstract

We introduce a corpus of short texts in Mandarin, in which quantified expressions figure prominently. We illustrate the significance of the corpus by examining the hypothesis (known as Huang's "coolness" hypothesis) that speakers of East Asian Languages tend to speak more briefly but less informatively than, for example, speakers of West-European languages. The corpus results from an elicitation experiment in which participants were asked to describe abstract visual scenes. We compare the resulting corpus, called MQTUNA, with an English corpus that was collected using the same experimental paradigm. The comparison reveals that some, though not all, aspects of quantifier use support the above-mentioned hypothesis. Implications of these findings for the generation of quantified noun phrases are discussed. MQTUNA is available at: https://github.com/a-quei/qtuna.

1 Introduction

Speakers trade-off clarity against brevity (Grice, 1975). It is often thought that speakers of East Asian languages handle this trade-off differently than those who speak Western European languages such as English (Newnham, 1971). This idea was elaborated in Huang (1984), when Huang borrowed a term from media studies, hypothesizing that Mandarin is "cooler" than English in that the intended meaning of Mandarin utterances depends more on context than that of their English counterparts; in other words, Mandarin speakers make their utterances shorter but less clear than English speakers. This "coolness" hypothesis is often worded imprecisely, conflating (a) matters that are built into the grammar of a language (e.g., whether it permits *number* to be left unspecified in a given sentence position), and (b) choices that speakers make from among the options that the grammar permits. Here we focus on the latter.

Studies of coolness have often focused on referring expressions (e.g., van Deemter et al. (2017); Chen et al. (2018); Chen and van Deemter (2020); Chen (2022)). The present paper focuses on *quantification*, as in the Quantified Expressions (QEs) "*All A are B*", "*Most A are B*", and so on. In a nutshell, we want to know whether Mandarin speakers use QEs less clearly, and more briefly, than English ones.

We report on an elicitation experiment, MQTUNA, inspired by the QTUNA experiment of Chen et al. (2019b, see §2). The experiment asks Mandarin speakers to produce sequences of QEs to describe abstract visual scenes. Sequences of QEs that are used to describe visual scenes are called Quantified Descriptions (QDs, Chen et al., 2019b). The MQTUNA corpus will enable researchers to investigate a wide range of questions about quantification in Mandarin. We illustrate this potential by comparing the corpus with the English QTUNA corpus from the perspective of coolness and we ask how our findings impact computational models of the production of QDs.

In sum, our contribution is two-fold:

1. We constructed, annotated and analysed the MQTUNA corpus;

2. We compared MQTUNA to QTUNA from the perspective of Huang's Coolness hypothesis.

2 QTUNA Experiment

A growing body of empirical work has studied how people understand and produce quantifiers (Moxey and Sanford, 1993; Szymanik and Zajenkowski, 2010; Grefenstette, 2013; Herbelot and Vecchi, 2015; Sorodoc et al., 2016). These studies have focused on a limited number of quantifiers (chiefly "*all*", "*most*", "*many*", and "*no*").

In Natural Language Generation (NLG), the QTUNA corpus was built to study how English
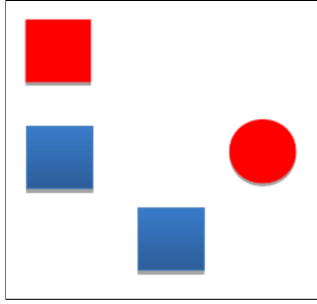
Figure 1: An example scene from QTUNA.

speakers use QDs to describe a visual scene (Figure 1). Participants were free to (1) describe a visual scene in whatever way they want, (2) use as many sentences as they choose, and (3) use any sentence pattern that they choose. For example, for the scene in Figure 1, a participant could say "*Half of the objects are blue squares. The other half are red objects. There is only one red circle.*". Given the domain contains four objects in no more than two shapes, this QD describes the scene completely and correctly. Participants were told that their descriptions should allow readers to reconstruct the scene *modulo* location. Each scene contains N objects (NB: N is defined as *domain size*), which is either a circle or a square and either blue or red. To test how domain size impacts the use of quantifiers, QTUNA experimented on 3 sizes, i.e., 4, 9, and 20.

Analysis of the resulting QTUNA corpus revealed that, as the domain size increase, (English) speakers (1) use more vague quantifiers (e.g., *most* and *few*); (2) use less complete QDs (NB: a QD is complete if the scene described is the only one *modulo* location that fits the description); (3) use more incorrect QDs (NB: a QD is incorrect if it is not true with respect to the scene); and (4) do not use longer QDs (measured in terms of the number of QEs).

3 Research Questions

Are the QTUNA findings true for MQTUNA? We are curious whether the above-mentioned findings about QTUNA (see §2) hold true for MQTUNA. We expected that domain size affects speakers of different languages in the same way, so these findings should hold for both corpora in the same way.

Are Mandarin QDs briefer and less clear than English QDs? "Coolness" says Mandarin speakers speak more briefly and less clearly than English speakers. We check this hypothesis by comparing QDs in QTUNA and MQTUNA.

Regarding brevity, we are curious about the length of QDs. If Mandarin QDs are briefer than English QDs, then we expect QDs in MQTUNA to contain less QEs than those in QTUNA.

Regarding clarity, if Mandarin speakers utter QDs in a less clear way, we expect to see more vague quantifiers in MQTUNA than in QTUNA and, more importantly, fewer logically complete QDs.

4 MQTUNA Experiment

We followed the same methodology as in the QTUNA experiment, re-using scenes of the QTUNA experiment, inheriting its experimental design, and translating its instructions participants.

4.1 Materials

To prepare materials for the MQTUNA experiment, we sampled scenes from QTUNA following two steps. First we eliminated all scenes all of whose objects share the same properties. For instance, we removed all scenes that can be described completely by a single QD like "*all objects are red circles*". Next, for each domain size (i.e., 4, 9, or 20), we randomly sampled 5 scenes from QTUNA. In the second step, to familiarise participants with the experiment, we added a practice situation that uses a $N = 4$ scene whose objects are the same.

For the instructions, we translated the instructions of QTUNA (Appendix A). More specifically, the instruction told subjects that (1) they should finish the experiment in limited time (i.e., 20 minutes); (2) their descriptions would then be used in a reader experiment where readers are asked to reconstruct the scenes; (3) they should not enumerate and not say where in the grid a particular object is located.

4.2 Design, Participants, and Procedure

Data from 31 participants were collected for domain sizes $N = 4$, 9 and 20 (N is the number of objects in the scene). See Appendix B for details about participants. Participants were asked to read the instruction first and to complete the experiment (16 situations) in one sitting.

4.3 The MQTUNA Corpus

The resulting MQTUNA corpus contains 465 valid QDs and 1175 QEs. There are 155 QDs for each domain size and there are 383, 386, and 406 QEs for $N = 4$, $N = 9$, and $N = 20$ respectively. Table 1 lists a number of examples QDs in MQTUNA.

N	Description
4	所有都是蓝色，方块是圆形三倍。 | *All objects are blue. The number of squares is triple that of circles.*
4	所有图形都是蓝色的。但是只有一个圆。 | *All objects are blue but there is only one circle.*
9	所有的圆圈是红色的。方块都是蓝色的。方块的数量少于圆圈的数量。 *All circles are red. All squares are blue. There are fewer squares than circles.*
9	方块是圆圈数量的三倍。全部为红色。 | *The number of squares is triple that of circles. All of them are red.*
20	图中红色蓝色方块圆球数量相差不大。 | *There is no big difference between the numbers of all combinations.*
20	一半红色，一半蓝色。红色方块比蓝色方块多。蓝色圆圈多于红色圆圈。 *Half of the objects are red, the other half of them are blue. There are more red squares than blue squares and more blue circles than red circles.*

Table 1: List of example descriptions from the MQTUNA corpus, with their annotations. N indicates domain size.

	$N = 4$	$N = 9$	$N = 20$
Quantified Description	155	155	155
Quantified Expression	383	386	406
Complete Description	122	19	5
Incomplete Description	33	136	150
Vague Quantifier	25	143	184
Wrong Description	7	14	30

Table 2: Frequencies of major QE types in the different subcorpora of MQTUNA.

We annotated the use of quantifiers in MQTUNA, viewing quantifiers that have the same meaning (e.g., "所有" ("suoyou", all) and "全部" ("quanbu", all) as identical. See Appendix C for a list of top-10 quantifiers and their usage in MQTUNA.

As for quantifier use, the quantifier "所有" (suǒyǒu; *all*) and "一半" (yíbàn; *half*) are two of the most frequent quantifiers. In the top-10 most frequent quantifiers of MQTUNA, 4 are vague, including "绝大多数" (*overwhelming majority*), "大多数" (*most*), "多数" (*most*), "少数" (*minority*). This is very different from QTUNA, where only 1 vague quantifier (i.e., *most*) is in top-10. Appendix C also presents lists of crisp and vague quantifiers.

5 Analysis

Focusing on the research questions of §3, we analyse the MQTUNA corpus (§5.1), and we compare MQTUNA with QTUNA (§5.2). We conclude with a few post-hoc observations (§5.3).

5.1 Analysing MQTUNA

To check whether the findings of QTUNA (§2) hold for MQTUNA, we annotated each QD with whether it is logically complete and whether it is correct with respect to the corresponding scene; we also annotated each QE with whether it uses a vague quantifier or not. Annotation details can be found in Appendix D. To avoid compromising the comparison between MQTUNA and QTUNA, we did not only annotate MQTUNA but we also re-annotated the QTUNA corpus, using the same annotators following the same set of principles. Table 2 charts the results.

Vagueness. We identified 57, 201, and 234 QEs that contain vague quantifiers out of 383, 386, and 406 QEs from the three sub-corpora, confirming that vagueness is more frequent with increasing domain size. This was confirmed by a binary logistic regression test ($p < .0001$, adjusted $p < .0001$[1]).

Completeness. We observed 33, 136, and 150 logically incomplete QDs from the three sub-corpora. A binary logic regression test confirms that there are more logically incomplete QDs in larger domains ($p < .0001$, adjusted $p < .0001$).

Correctness. The 3 subcorpora contained 7, 14, and 30 wrong QDs, so more incorrect QDs are used in larger domains ($p < .0001$, adjusted $p < .0001$) using a binary logic regression test.

Length. QDs in larger domains in MQTUNA contain more QEs than those in smaller domains. We computed the Pearson correlation between the domain size and the QD length. After Bonferroni correction, the difference fell just short of significance ($p = 0.1025$, adjusted $p = 0.615$).

In a nutshell, all findings of QTUNA are also true for MQTUNA.

5.2 MQTUNA and QTUNA: Initial Comparison

To compare QDs in Mandarin and English, we considered all the scenes in MQTUNA and re-annotated them using the same annotators in §5.1.

---

[1]The p-value was adjusted by Bonferroni correction

	QTUNA		MQTUNA		
N	C	I	C	I	p-value
4	298	32	122	33	$p < .001$
9	77	113	19	136	$p < .0001$
20	4	186	5	155	$p = .5$
all	379	331	146	319	$p < .0001$

Table 3: Numbers of complete (C) and incomplete (I) QEs in QTUNA and MQTUNA. N is domain size.

Brevity. We compared the length of QDs in QTUNA and MQTUNA and found that QDs in MQTUNA are longer than those in QTUNA in every sub-corpus. This rejects our hypothesis that Mandarin speakers prefer brevity and, thus, produce shorter QDs than English speakers.

Completeness. Table 3 reports the number of logically complete QDs in QTUNA and MQTUNA, respectively. 379 out of 710 QDs in QTUNA are logically complete while 146 out of 465 QDs in MQTUNA are complete. Using a Chi-squared test, this confirms that there are more complete QDs in QTUNA than in MQTUNA ($\chi^2(2, N = 1175) = 54.93, p < .0001$, adjusted $p < .0001$). Mandarin speakers produce longer but less logically complete QDs. Interestingly, if we look into more details (see Table 3), the difference only exists in domain sizes 4 and 9. We suspect that both English and Mandarin speakers find it hard to come up with a logically complete QD if the domain size is large.

Vagueness. In QTUNA, 222 of the 1342 QEs were vague whereas, in MQTUNA, 352 of the 1175 QEs were vague. A Chi-squared test confirms that Mandarin speakers used more vague quantifiers than English speakers ($\chi^2(2, N = 2517) = 64.04, p < .0001$, adjusted $p < .0001$).

5.3 Post-hoc Observations

Surface Forms. We observed that QEs in MQTUNA are generally realised in three kinds of forms: (1) "Q A 是 B" ("*Q A are B*"), where "Q" is a quantifier, for example, "大部分 A 是 B" ("*most A are B*"); (2) "A 中 Q 是 B" ("*in A, Q are B*"); and （3）"B 在 A 中 占 Q" ("*B takes up Q of A*").

A-Drop. Akin to the previous findings that pronouns and nouns are often dropped in Mandarin NPs (Huang, 1984; Osborne and Liang, 2015), we found that nouns that take up A positions in the above forms are also often dropped (henceforth, *A-drop*), for example, saying "B 占 Q" ("*B takes up Q*"). In MQTUNA, we found 304 out of 1175 QEs (approximately 25.87%).

Plurality. van der Auwera and Baoill (1998) pointed out that Mandarin briefer in that plurality is often not expressed explicitly. Consistent with this, we found that in MQTUNA, numbers are rare. This makes a QE in Mandarin sometimes less informative than an English QE, Mandarin QDs are less likely to be logical complete. For example, Mandarin QE "图片中有红色方块" could mean "there are red squares" or "there is a red square".

6 Discussion

We have presented and analysed the MQTUNA corpus of quantifier use in Mandarin.

Coolness. We assessed the coolness hypothesis by analysing MQTUNA and comparing QTUNA and MQTUNA. As for the brevity of QDs, we found both evidence (i.e., Mandarin speakers often performed A-drop and expressed plurality implicitly) and counter-evidence (i.e., Mandarin speakers uttered longer QDs than English speakers).

As for the clarity of QDs, we confirmed that the Mandarin corpus (MQTUNA) contains significantly more *incomplete* QDs and *vague* quantifiers than its English counterpart (QTUNA).

Generating QDs. Chen et al. (2019a) proposed algorithms for generating QDs (QDG algorithms). Let us list issues to be heeded when building QDG algorithms for Mandarin.

First, plurality plays an important role in the QDG Algorithms of Chen et al. (2019a). If these algorithms are to be adapted to Mandarin, then they should first "decide" whether to realise the plurality of a QE explicitly, since this will influence how much information the QD should express in other ways. Second, modelling the meaning of vague quantifiers is vital for generating human-like QDs. Since Mandarin speakers use vague quantifiers more frequently than English speakers, Mandarin QDG needs to handle a larger number of vague quantifiers and capture nuances between them, which is a difficult and data-intensive challenge. Lastly, QD surface realisation in Mandarin needs to handle more syntactic variations than current QDG algorithms are capable of, because (1) a QE can be realised in multiple possible forms (see §5.3); (2) A-drop frequently happens; (3) Plurality can be expressed implicitly or explicitly.

Future Work. Our comparison between Mandarin and English was based on two corpora, QTUNA and MQTUNA, that were collected using elicitation experiments that were conducted following the same experimental paradigm, and using very similar sets of stimuli. Yet, *language* may not have been the only difference between these experiments; participants in QTUNA and MQTUNA are also likely to differ in terms of their *cultural background*, and possibly in terms of other variables, such as their education; There is no absolute guarantee that all our annotations are correct. To create an even playing field between the two corpora, we asked our annotators to re-annotate QTUNA. But although our annotator were native speakers of Chinese, they were merely fluent (not native) in English, which may have caused a difference in the way both corpora were annotated. In future, it would be interesting to conduct even more tightly controlled experiments to tease apart the variable of language use from such possibly confounding variables.

Finally, our experiment has looked at a wide range of quantifiers. We also plan experiments that zoom in on specific subsets, such as the different ways in which variants of the quantifier "*most*" can be expressed (cf., Wang and Piao (2007) and §4.3).

References

Guanyi Chen. 2022. *Computational Generation of Chinese Noun Phrases*. Ph.D. thesis, Utrecht University.

Guanyi Chen and Kees van Deemter. 2020. Lessons from computational modelling of reference production in Mandarin and English. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 263–272, Dublin, Ireland. Association for Computational Linguistics.

Guanyi Chen, Kees van Deemter, and Chenghua Lin. 2018. Modelling pro-drop with the rational speech acts model. In *Proceedings of the 11th International Conference on Natural Language Generation*, pages 159–164, Tilburg University, The Netherlands. Association for Computational Linguistics.

Guanyi Chen, Kees van Deemter, and Chenghua Lin. 2019a. Generating quantified descriptions of abstract visual scenes. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 529–539, Tokyo, Japan. Association for Computational Linguistics.

Guanyi Chen, Kees van Deemter, Silvia Pagliaro, Louk Smalbil, and Chenghua Lin. 2019b. QTUNA: A corpus for understanding how speakers use quantification. In *Proceedings of the 12th International Conference on Natural Language Generation*, pages 124–129, Tokyo, Japan. Association for Computational Linguistics.

Edward Grefenstette. 2013. Towards a formal distributional semantics: Simulating logical calculi with tensors. In *Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity*, pages 1–10, Atlanta, Georgia, USA. Association for Computational Linguistics.

Herbert P Grice. 1975. Logic and conversation. In *Speech acts*, pages 41–58. Brill.

Aurélie Herbelot and Eva Maria Vecchi. 2015. Building a shared world: mapping distributional to model-theoretic semantic spaces. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 22–32, Lisbon, Portugal. Association for Computational Linguistics.

C-T James Huang. 1984. On the distribution and reference of empty pronouns. *Linguistic inquiry*, pages 531–574.

Linda M Moxey and Anthony J Sanford. 1993. *Communicating quantities: A psychological perspective.* Lawrence Erlbaum Associates, Inc.

Richard Newnham. 1971. *About Chinese*. Penguin Books Ltd.

Timothy Osborne and Junying Liang. 2015. A survey of ellipsis in chinese. In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 271–280.

Ionut Sorodoc, Angeliki Lazaridou, Gemma Boleda, Aurélie Herbelot, Sandro Pezzelle, and Raffaella Bernardi. 2016. "look, some green circles!": Learning to quantify from images. In *Proceedings of the 5th Workshop on Vision and Language*, pages 75–79, Berlin, Germany. Association for Computational Linguistics.

Jakub Szymanik and Marcin Zajenkowski. 2010. Comprehension of simple quantifiers: Empirical evaluation of a computational model. *Cognitive Science*, 34(3):521–532.

Kees van Deemter, Le Sun, Rint Sybesma, Xiao Li, Bo Chen, and Muyun Yang. 2017. Investigating the content and form of referring expressions in Mandarin: introducing the mtuna corpus. In *Proceedings of the 10th International Conference on Natural Language Generation*, pages 213–217, Santiago de Compostela, Spain. Association for Computational Linguistics.

Johan van der Auwera and Dónall Ó Baoill. 1998. *Adverbial constructions in the languages of Europe*, volume 3. Walter de Gruyter.

Amy Y Wang and Scott Piao. 2007. Translating vagueness? a study on translations of vague quantifiers in an english-chinese parallel corpus. In *Proceedings of the Corpus Linguistics Conference*.

A Instruction

您好，我们最近的研究关注于人描述物体集合的方法。为此，我们设计了一个小实验。在这个实验中，我们将给您展示一系列图片。在每张图片中，您将看到一定数量（*16个*）的图形。在看到每张图片后，我们需要您写一句或几句语法正确的中文句子。请注意：
We're interested in understanding how people describe sets of objects. To find out, we're doing a small experiment, in which we'll show you a number of situations in which some (16) objects are displayed on a screen. We'd like you to describe each situation in one or more grammatically correct Mandarin sentences.

1 您将在有限的时间（*20分钟*）内完成整个实验。 *The experiment should take you less than 20 minutes.*

2 根据您写的描述，后续实验中的被试者会用它来在有限时间（总共*20分钟*）内重构图片。"重构"的在这里仅表示图片中每种图形数量。因此在您的表述中，您不必描述每个图形在图片中的位置（例如：上方，在中间）。 *Based on your description, a reader will try to "reconstruct" the situation. We use the word "reconstruct" loosely here, because the only thing that matters is the different types of objects that the sheet contains. Therefore, please do not say *where* in the grid a particular object is located (e.g., "top left", "in the middle", "on the diagonal").*

3 每个图形可能是方形也可能是圆形，可能是红色也可能是蓝色。后续负责重构的被试者也知晓这个信息。负责重构的被试者同时还知晓图片中图形的数量。这些信息都会被用在重构当中。 *Each object is a circle or a square, and either red or blue. Your reader knows this.*

4 请不要"枚举"图片中的图形，例如：图片中有一个红色的圆圈，两个蓝色的圆圈，和三个蓝色的方块。 *Please do not "enumerate" the different types of objects. For example, do not say "There is a red circle, two blue circles, and ...".*

以下是几个例子*:*
Here are some Example: (...)

Figure 2: The sketch of the instruction of MQTUNA.

B Participants

All of our participants are Mandarin native speakers. 21 subjects are undergraduate students in computer science from the Utrecht University. Each of the rest at least has a bachelor degree in any of computer science, statistics, and management. 11 subjects are female and 20 subjects are male.

C Quantifiers in MQTUNA

Table 4 enumerates the top-10 quantifiers and their usage in MQTUNA. In what follows, we provide a list of vague quantifiers and a list of crisp quantifiers in MQTUNA.

- Crisp Quantifiers: 所有 (*all*), 只有 (*only*), 比...多... (*more*), 倍 (*times*), 除了...都是... (*all...except...*), 有 (*there is*), 多于n倍 (*more than n times*), 少于n倍 (*less than n times*), 各半 (*half...the other half...*), 相同 (*same as*), 一半 (*half*), 不同 (*different amount of*), 一半以上 (*more than half*), 没有 (*no*), 少于 (*less than*), 所有组合 (*all possible combinations*);

- Vague Quantifiers: 大部分 (*most*), 小部分 (*a small part of*), 绝大部分 (*overwhelming majority*), 除了...大多数... (*most...except...*), 少量的 (*a few*), 远多于 (*way more than*), 极少数 (*a very few*), 多一点 (*slightly more than*), 多不少 (*greatly more than*), 相近 (*close to each other*), 基本都是 (*almost all*), 略少 (*a bit less*), 略多 (*a bit more*), 大约各半 (*approximately half ... the rest ...*), 基本相同 (*almost the same*), 多一些 (*several more*), 多好几倍 (*several times more*), 多得多 (*much more*), n倍多一点 (*slightly more than n times*), n倍少一点 (*slightly less than n times*), 大约一半 (*approximately half*), 少数 (*minority*).

D Annotating MQTUNA

We asked our annotator to annotate logical completeness, correctness and vagueness based on the following principles:

1. Logical Completeness: we asked our annotator whether s/he can fully recover the scene given a QD. For example, for a scene with 3 red circles and 1 blue square, one could say "*Most objects are red circles and there is only one blue square.*" Though s/he uses a vague quantifier "most", we still can infer that, given domain size 4, "most objects" means 3 objects, and, therefore, this QD is logically complete. However, for a scene with 8 red circles and 1 blue circle, one could say "*All objects are circles and almost all of them are 8.*" Though using "almost all" to describe "8 out of 9" is definitely correct, it does not necessarily mean "8 out of 9" but possibly mean "7 out of 9". Therefore, this QD is not logically complete;

2. Correctness: we asked our annotator to annotate a QD as "incorrect" if and only if the QD contains definitely incorrect information, for example, saying a "red object" blue or describing a scene with 3 red squares and 1 blue square as "*half of the objects are red*";

3. Vagueness: our annotator decided whether a QE uses a vague quantifier based on the vague quantifier list in Appendix C.

Notation	English	Surface Form(s)	Example Quantified Expression(s)	Frequency			
				N=4	N=9	N=20	Total
所有	all	(所有)...都..., (全部)...都...	(全部)A都是B / A中(全部)都是B *All A are B*	100	127	53	280
一半	half	一半, 百分之五十	一半A是B / A中的一半是B / B在A中占一半 *Half A are B*	101	19	28	148
相同	equal	数量相同, 一样多, 个数一样	A与B数量相同 *There is an equally number of A and B*	59	11	29	99
绝大多数	overwhelming majority	绝大部分, 绝大多数	A中绝大多数是B / 绝大多数A是B / B在A中占绝大多数 *Almost A are B*	7	50	37	94
各半	half ... rest ...	各半, 一半一半, 一半...另一半...	BC在A中各半 / A中BC各半 / 一半的A是B, 另一半是C *Half of A are B, the other half of A are C*	60	6	24	90
比-多	more	比...多	A比B多 *A more B*	10	28	48	96
大多数	most	大多数, 大部分	A中大多数是B / 大多数A是B / B在A中占大多数 *Most A are B*	7	35	33	75
少数	minority	少数, 少部分	A中少数是B / 少数A是B / B在A中占少数 *Minority of A are B*	5	31	24	60
有	exist	有, 存在	图片中有A (*There are A in the scene*)	4	12	18	34
多数	most	多数	A中多数是B / 多数A是B / B在A中占多数 *Most A are B*	5	4	20	29

Table 4: Top-10 most frequently occurring quantifiers with their English translation and Mandarin examples as well as frequencies in the three MQTUNA sub-corpora.