

Definiteness: from L1 Mandarin to Mandarin L2 English*

Bert Le Bruyn and Xiaoli Dong

Utrecht Institute of Linguistics, Utrecht University, The Netherlands
b.s.w.lebruyne@uu.nl, x.dong@uu.nl

1 Introduction

In this paper we present a preview of our study on the L2 acquisition of English definites by L1 Mandarin speakers, highlighting those aspects that are relevant to the current theoretical debate on definiteness. Section 2 zooms in on the learners' L1 and argues that – next to standard definite uses of bare nouns – it also has a construction that marks non-maximal familiar definiteness. Section 3 zooms in on the learners' L2, argues that they don't confuse definiteness/specificity and locates the main acquisition problem in the constraints that govern accommodation.

Our pretheoretical working definitions are as follows: we assume maximality is to be understood as exhaustivity at the level of the model and that familiarity implies previous mention in the discourse. Both the model and the discourse can be contextually restricted (see e.g. Brisson, 1998; Walker et al., 1998).

2 Mandarin

In the recent literature on cross-linguistic variation in the domain of definiteness, two papers stand out, *viz.* Schwarz (2009) and Arkoh and Matthewson (2013). Both argue that there are languages with a definite article that marks non-maximal familiar definiteness. Continuing the work of Yang (2005) and Partee (2006), we argue that Mandarin is relevant in this discussion as well. Next to the definiteness conveyed by bare nouns (see e.g. Cheng and Sybesma, 1999, 2012), it also has a construction that marks non-maximal familiar definiteness. The study of non-maximal familiar expressions is important for familiarity- as well as for uniqueness-based theories of definiteness.

The construction we are referring to consists of the sequence *modifier + numeral + classifier + noun*, henceforth *MNCN*. The modifier can be – among other things – a possessive or a relative clause:

- (1) Zhangsan de san ben shu
John DE three CL book
'John's three books'
- (2) wan tiaosan de liang ge xuesheng
play parachute DE two CL student
'the two skydiving students'

*This paper presents a preview of the data in Le Bruyn and Dong (prep). We thank the semantics crowds in Utrecht and Leiden, Jeannette Schaeffer, our research assistants Xinyuan Wang and Wenzhu Xuan as well as all our participants. Special thanks to Yunhua Hu and his colleagues for helping us with our experiments at the Zhejiang Ruian High School and our family and friends in China for their hospitality. Another special thanks to Anna Volkova for preparing the L^AT_EX version of this paper. The first author furthermore gratefully acknowledges the support of NWO, grant 275-80-006.

Previous analyses have qualified the MNCN as specific, referential, not non-definite (Huang, 1982), familiar, non-maximal definite (Yang, 2005; Partee, 2006), specific indefinite (Sio, 2006; Zhang, 2006), and familiar, maximal definite (Hall, 2015). We ran two experiments to assess the exact nature of the construction.

2.1 Experiment 1: familiarity

The aim of the first experiment was to establish under which circumstances the use of the MNCN would be licit. We distinguished between three conditions: one in which the speaker could not identify the referent, one in which the speaker could identify the referent and one in which the speaker referred to a previously introduced referent.

The different items were all integrated into one overarching story about a guard in an elite sports school in China. The school had 6 students and both the girls and the boys could easily be distinguished during the day but happened to have the same silhouettes, thus making identification during the night impossible. The day vs. night setting was the crucial difference between the first two conditions.

(3) The students



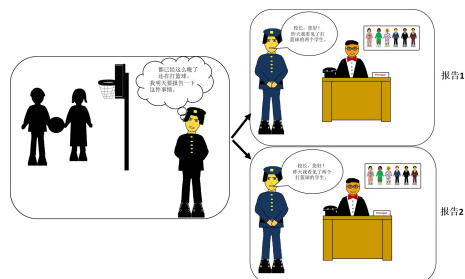
Participants (n=15) were asked to play the guard who reported on what he had seen to the principal. Their input was a visualization of what the guard had seen and they had to select a description with an MNCN or one in which the modifier was put between the classifier and the noun. These descriptions were put into text balloons in pictures representing the guard reporting to the principal.

The third condition involved objects that had been stolen from the principal's office and that the principal had asked the guard to locate (e.g. sports trophies). Given the importance of these objects, the principal had asked the guard to call him as soon as he found them. The MNCN and non-MNCN reports of the discovery of the stolen items were put into text balloons in pictures representing the guard calling the principal.¹

(4) to (6) give an impression of each of the respective conditions. In (4) we also provide a written out version of the text balloons, a gloss and a translation.

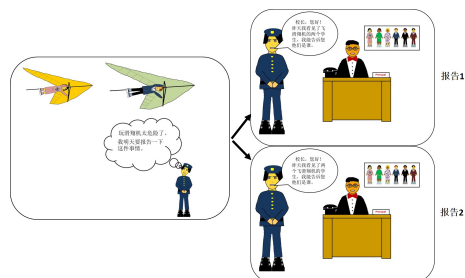
¹We also designed a version of this experiment in which the third condition involved students instead of objects. The results were comparable but the overarching story was slightly more elaborate. For reasons of space we only report on the experiment with objects.

(4) Non-identifiable referents

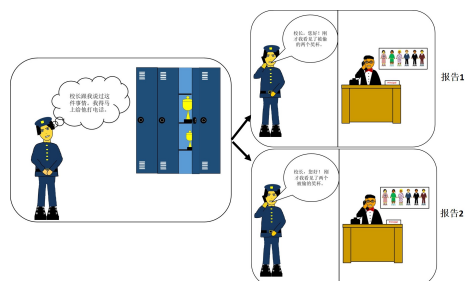


- a. Dou yijing zhemewan le, hai zai da lanqiu, womingtian yao baogao
PRT already so late PRT, still PRT play basketball, I tomorrow will report
yixia zhe jian shiqing.
quickly this CL affair
'It's already so late and they're still playing basketball, I'll briefly report about this
affair tomorrow.'
- b. Xiaozhang, ninhao! Zuotian wo kanjian le da lanqiu de liang ge xuesheng.
Principal, hello! Yesterday I see LE play basketball DE two CL student
'Principal, hello! Yesterday I saw two students playing basketball.'
- c. Xiaozhang, ninhao! Zuotian wo kanjian le liang ge da lanqiu de xuesheng.
Principal, hello! Yesterday I see LE two CL play basketball DE student
'Principal, hello! Yesterday I saw two students playing basketball.'

(5) Identifiable referents



(6) Previously introduced referents



We ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that the familiar condition was the only one that was significantly different from the other two in allowing the use of the MNCN (familiar/simple indefinite: $t(177)=5.278$; $p<.000$ | familiar/ specific indefinite $t(177)=4.987$; $p<.000$). The interpretation we give of these results is that familiarity is a necessary condition for using the MNCN in Mandarin.

2.2 Experiment 2: maximality

The second experiment assessed maximality. The rationale behind it was that definites that allow for maximal readings will favor those in out of the blue contexts, thus making sentences like (7) sound as inconsistent as sentences like (8), involving *all*:

(7) ?I stole the two pens of Mary's, so now she only has one left.

(8) ?I stole all of Jacky's books, so now she only has one left.

An experiment with English native speakers ($n=15$) recruited via Crowdfunder confirmed that (7) and (8) are not significantly different from each other and that both differ from sentences with indefinites as in (9):

(9) Harry smoked 5 cigarettes, so now he only has 3 left.

Examples of the Mandarin items corresponding to (7) and (8) are given in (10) and (11), (10) involving the MNCN as the counterpart of the English definite, and (11) containing *suoyou* as the counterpart of English *all*.

(10) Wo chi le Xiaowang de san ge jidan, suoyi ta xianzai zhiyou san ge le.
I eat LE Xiaowang DE three CL chicken egg, so he now only CL LE

(11) Xiaoding chi le suoyou de dangao, suoyi xianzai zhiyou san ge dangao le.
Xiaoding eat LE all DE cake, so now only three CL cake LE

Next to items like (10) that have a non-relational noun (*egg*) and a possessor as modifier (*Xiaowang de*), we also included MNCNs with relational nouns and relative clauses as modifiers, giving rise to three subtypes of MNCN conditions (possessor modifier + relational noun; possessor modifier + non-relational noun; relative clause modifier + non-relational noun).

Participants ($n=17$) were asked to assess the consistency of the sentences we presented them with, focusing on the link between the first and the second part. They received three answer options (compatible, possibly compatible and incompatible). In our analysis we reduced these to two options (compatible/possibly compatible vs. incompatible).

We ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that all MNCN conditions were judged significantly more consistent than the all condition ($t(183)=6.615$ | 6.801 | 6.897 ; $p<.000$) but did not differ from each other. The interpretation we give of these results is that maximality is not part of the interpretation of the MNCN in Mandarin.

2.3 The MNCN: conclusion

Based on the two experiments, we conclude that familiarity is a necessary condition for the use of the MNCN and that maximality is not part of its interpretation. The only analysis in the literature that is compatible with these data is that of Yang (2005) and Partee (2006), who assume the MNCN expresses non-maximal familiar definiteness.

3 Mandarin L2 English

In the past decade the literature on the L2 acquisition of definiteness by L1 speakers of articleless languages (like Korean, Russian, Mandarin, Polish, etc.) has almost exclusively focused on the complicating role of specificity in this process. The consensus seems to be that L1 speakers of articleless languages – independently of proficiency – often have a hard time deciding what the exact semantics of the definite article is, fluctuating between definiteness and specificity. We argue that specificity is not at play (3.1) and that the main acquisition problem lies in the acquisition of maximality (3.2).

3.1 Assessing the influence of specificity

In order to assess the influence of specificity on the L2 acquisition of definites, we designed a new experimental paradigm. The standard paradigm used in the literature is that of Ionin et al. (2004) but has evolved over the years (compare e.g. Ionin et al. (2004) and Ko et al. (2010)) and the construct *specificity* in the experiment has never been fully validated (see Ionin (2006) for discussion).²

Ionin (2006) takes the type of specificity that comes into play in the L2 acquisition of definiteness to be identical to the one expressed by indefinite *this* in English:

- (12) If you did get in any trouble, there was **this man**, Joe, and he was a really big man, he was a black-belt in karate and everything, and ... (example taken from the BNC)

Ionin describes the use of *this* in examples like (12) as signaling that the speaker intends to refer to exactly one individual who is moreover noteworthy in some way or other. We operationalized these intuitions in various ways and ran a series of experiments with native speakers of English we recruited online via Crowdfunder. The crucial criterion for success was a significant difference between non-specific and specific contexts in welcoming the use of *this* rather than *a*. We succeeded in meeting this criterion by playing with a foreground/background distinction as in (13) and (14) where indefinite *this* was judged more appropriate for the foregrounded referent of *this boy* in (13) than for the backgrounded referent of *this brick* in (14).

- (13) Have I already told you my favorite soccer story? Well, one day I saw **this boy** playing soccer in our street... All of a sudden he jumps up, makes an amazing salto and makes the ball hit a can. It turned out he was the son of a famous soccer player.
- (14) Have I already told you about the scariest moment of my life? Well, one day I saw a girl on top of a building... All of a sudden, she starts to dance, slips on **this brick** and falls off the building! Fortunately she landed on some cardboard boxes and didnt get hurt...

The specific question we asked the participants (n=15) was whether the use of *this* was felicitous in the given contexts or whether they would prefer to replace it by *a*.

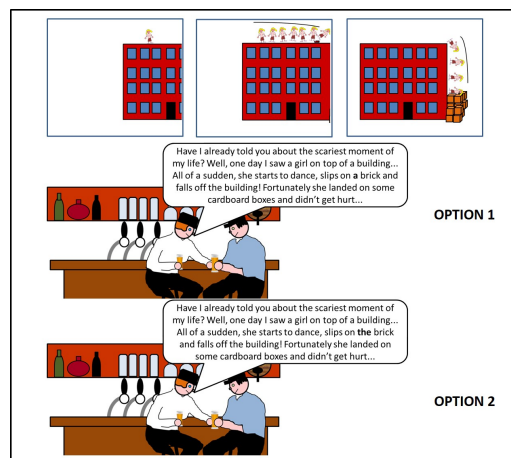
We ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that the foregrounded uses of *this* were significantly more felicitous than the backgrounded uses ($t(118)=2.326$; $p=.022$).

With a successful operationalization of specificity in place, we set up an L2 experiment in which we included non-specific and specific indefinite conditions as well as a familiar definite condition. In each item we presented two versions of a small story, one with an indefinite article and one with a definite article. We asked Mandarin L2 learners of English (n=19)

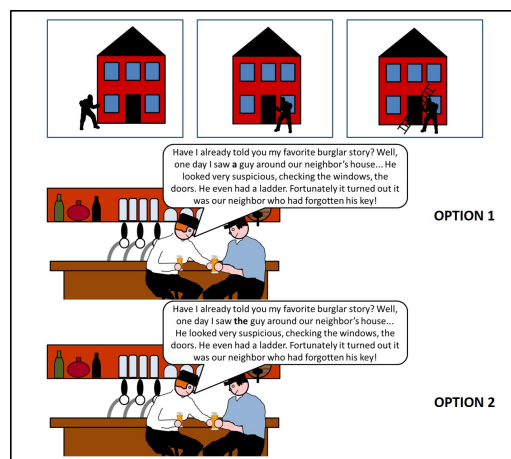
²See the end of section 3.2 for some comparison between the two paradigms.

recruited through Crowdfunder to choose the story that sounded most natural to them.³ To help participants, we visualized the discourse setting of the story (the story was told by one guy to another guy in a pub) as well as the story itself. (15) to (17) illustrate the different conditions:

(15) Non-specific indefinite

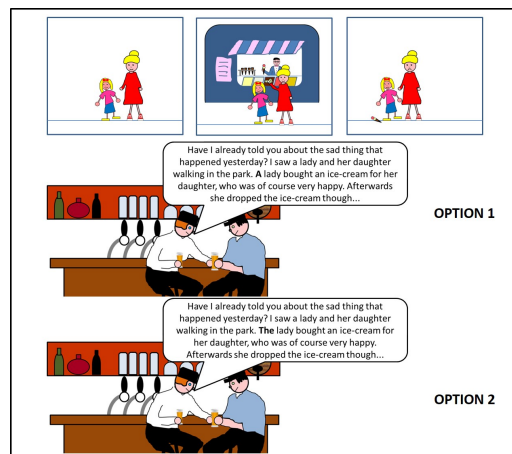


(16) Specific indefinite



³We recruited 30 participants in China and Taiwan (not including Hong-Kong) that were level 2 or 3 workers in Crowdfunder and were moreover identified as speakers of Chinese by Crowdfunder. We asked each participant to indicate whether Putonghua ('standard Mandarin') was their mother tongue and excluded the participants (n=5) who indicated that it wasn't. The survey furthermore included 8 items of the Michigan Quick Placement Test turned into a grammaticality judgement task which served as test questions in Crowdfunder and furthermore allowed us to get a rough idea of the participants' proficiency in English. We excluded participants (n=6) who didn't get at least 5 items right. This exclusion had no effect on the statistical results.

(17) (Familiar) definite



If specificity were to play the role it has been claimed to play in the literature, we would expect the definite condition and the specific indefinite condition to be eliciting significantly more definites than the non-specific indefinite condition. To check this prediction, we ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that the non-specific indefinite condition elicited significantly more definites than the specific indefinite condition ($t(117)=3.096$; $p=.002$) and that the definite condition still elicited significantly more definites than the non-specific indefinite condition ($t(117)=6.152$; $p<.000$). This effect is not the one we would expect if Mandarin L2 English learners were to fluctuate between a specific and a definite semantics for the definite article. It however receives a straightforward – be it *post hoc* – explanation if we assume Mandarin L2 English learners assume the definite article marks familiarity. The higher rate of definites in the non-specific condition then follows from the well-known connection between backgrounded and given information.

3.2 The acquisition of maximality

In 3.1 we established that specificity is not interfering with the acquisition of definiteness. We however know that Mandarin L2 English learners do overgenerate definites (see e.g. Tryzna, 2009; Yang and Ionin, 2009b, etc.). We propose an explanation of these data couched in theories of definiteness.⁴

We assume Mandarin L2 English learners start with an interlanguage grammar in which definites are familiar but not maximal. Support for this comes from three experiments. The first is the one we presented in 3.1, the other two are briefly summarized here:

- An exploratory production experiment (a culturally modified version of the paradigm used by Schaeffer and Matthewson 2005):

We tested 2nd grade high school students from the Zhejiang Ruian High School ($n=16$) and found that they performed at ceiling for the production of definites in the familiar

⁴A comparison with Yang and Ionin (2009a) will appear in Le Bruyn and Dong (prep).

condition. This indicates that Mandarin L2 learners of English associate definites with familiarity.

- A felicity judgement experiment (identical to the one we discussed for native English speakers in section 2.2):

We tested students from the Beijing International Studies University (n=20) and found that they differed from native English speakers in finding sentences like *I stole the two pens of Mary's, so now she only has one left* significantly more consistent than sentences like *I stole all of Jacky's books, so now she only has one left*. At the same time they however also judged sentences with *the* less consistent than sentences like *Harry smoked 5 cigarettes, so now he only has 3 left*.

We ran a mixed effects model with item and participant as random factors and found a significant effect of condition. Pairwise comparisons showed that the definite condition was significantly different from the *all* condition ($t(237)=12.820$; $p<.000$) and from the indefinite condition ($t(237)=3.184$; $p=.002$). These data indicate that Mandarin L2 learners of English differ from native speakers in not always associating definites with maximality.

In the process of acquisition, Mandarin L2 English learners are confronted with maximality as another ingredient of definiteness and pick up on this. Support for this comes from one experimental study and one corpus study:

- An exploratory forced choice elicitation task (an experiment including part of the items from Ko et al. 2010):

We tested 2nd grade high school students from the Zhejiang Ruian High School (n=35) and found that they performed almost at ceiling for the production of lexically unique referents (*the capital of North Dakota, the president of our university*) as well as classical bridging definites (*the cover* (of a previously mentioned book), *the screen* (of a previously mentioned laptop)). These data indicate that Mandarin L2 English learners – even though they don't establish a one-to-one correspondence between definites and maximality – are sensitive to maximality for the definite article.

- A study of the Portsmouth learner corpus:

We had two native speakers error tag 500 random occurrences of the definite article in texts produced by Mandarin L2 English learners (kappa: 0.82). The native speakers agreed on 58 problematic uses of *the*, several of which involved overproduction of the definite for kind reference. Given that these cases don't occur in the L2 and cannot be due to L1 transfer either, they are most likely cases in which the maximality semantics of the definite article was (erroneously) generalized to kind reference.

The acquisition problem that presents itself in moving from a familiarity to a familiarity + maximality semantics of the definite article lies in merging a familiarity and a maximality view on definiteness. The contexts that pose most problems are those in which nouns are used that are not lexically unique but come with sufficient modification to warrant a contextually unique interpretation. The felicity of a definite in these contexts depends on the availability of accommodation, a process whose conditions seem to be well understood by native speakers but pose a challenge for semanticists working on definiteness as well as for L2 learners with an articleless L1. An example is given in (18):

(18) Phone conversation

Jennifer: Hello, Helen? This is Jennifer!
Helen: Hi, Jennifer! It's wonderful to hear from you. I suppose you want to talk to my sister?
Jennifer: Yes, I haven't spoken to her in years! I'd like to talk to her now if possible.
Helen: I'm very sorry, but she doesn't have time to talk right now. She is meeting with a **very important client from Seattle. He is quite rich, and she really wants to get his business for our company!** She'll call you back later.

Chances are real that the company Jennifer is working for doesn't have more than one important client in Seattle. This means that – in principle – a contextually unique interpretation should be viable. The particular way of phrasing chosen in (18) however proscribes accommodation for native speakers and forces the use of *a*. Some slight variations would however felicitously host a definite:

- (19) She's meeting with the very important client who arrived from Seattle.
She's meeting with the very important client she met in Seattle.
She's meeting with the very important client from across the street.

The challenge for definiteness theories as well as for L2 learners with an articleless L1 is to master the subtle differences between (18) and (19).

The choice of (18) is not accidental. It's a specific indefinite item from one of the more recent test batteries of Ionin and colleagues (Ko et al., 2010). One of the most striking differences between their specific indefinite items and ours is that we have operationalized specificity by making the corresponding referents into the protagonists of a story that – at least according to the speaker – was worth telling. For Ionin and colleagues, specificity is operationalized as the enumeration of noteworthy properties. Given that these easily give rise to the impression that a contextually unique reading is targeted, it doesn't come as a surprise that they elicit overgeneration of definites by L2 learners with an incomplete mastery of the constraints on accommodation.

4 Conclusion

In this paper, we have given a preview of our work on the L2 acquisition of English definites by L1 Mandarin speakers. We have highlighted those aspects that are of relevance for current discussions about the nature of definiteness, in particular the relation between maximality and familiarity.

References

- Arkoh, R. and Matthewson, L. (2013). A familiar definite article in Akan. *Lingua*, 123:1–30.
Brisson, C. (1998). *Distributivity, maximality, and floating quantifiers*. PhD thesis, Rutgers, The State University of New Jersey.

- Cheng, L. and Sybesma, R. (1999). Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry*, 30(4):509–542.
- Cheng, L. and Sybesma, R. (2012). Classifiers and DP. *Linguistic Inquiry*, 43(4):634–650.
- Hall, D. (2015). Licensing definiteness. Ms. Queen Mary University London.
- Huang, C. T. (1982). *Logical relations in Chinese and the theory of grammar*. PhD thesis, MIT.
- Ionin, T. (2006). *This* is definitely specific: specificity and definiteness in article systems. *Natural Language Semantics*, 14(2):175–234.
- Ionin, T., Ko, H., and Wexler, K. (2004). Article semantics in L2 acquisition: The role of specificity. *Language Acquisition*, 12(1):3–69.
- Ko, H., Ionin, T., and Wexler, K. (2010). The role of presuppositionality in the second language acquisition of English articles. *Linguistic Inquiry*, 41(2):213–254.
- Le Bruyn, B. and Dong, X. (prep). The L2 acquisition of English definites by L1 Mandarin speakers. In preparation. Utrecht institute of Linguistics.
- Partee, B. H. (2006). A note on Mandarin possessives, demonstratives, and definiteness. In Birner, B. J. and Ward, G., editors, *Drawing the boundaries of meaning: Neo-Gricean studies in pragmatics and semantics in honor of Laurence R. Horn*, pages 263–280. Amsterdam: John Benjamins.
- Schaeffer, J. and Matthewson, L. (2005). Grammar and pragmatics in the acquisition of article systems. *Natural Language & Linguistic Theory*, 23:53–101.
- Schwarz, F. (2009). *Two types of definites in natural language*. PhD thesis, UMass, Amherst.
- Sio, J. (2006). *Modification and reference in the Chinese nominal*. PhD thesis, Leiden University.
- Tryzna, M. (2009). Questioning the validity of the article choice parameter and the fluctuation hypothesis. In del Pilar García Mayo, M. and Hawkins, R., editors, *Second language acquisition of articles*, pages 67–86. Amsterdam: John Benjamins.
- Walker, M., Joshi, A., and Prince, E. (1998). *Centering theory in discourse*. Oxford University Press.
- Yang, H. (2005). *Plurality and modification in Mandarin nominal phrases*. PhD thesis, University of Texas at Austin.
- Yang, M. and Ionin, T. (2009a). L2 English articles and the computation of uniqueness. In *Proceedings of the 3rd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, pages 325–335. Connecticut, US.
- Yang, M. and Ionin, T. (2009b). Specificity and speaker knowledge in the second language acquisition of English articles. Ms. Guangdong University of Foreign Studies and University of Illinois at Urbana-Champaign.
- Zhang, N. (2006). Representing specificity by the internal order of indefinites. *Linguistics*, 44(1):1–21.