Syntactic Cues to Individuation in Mandarin Chinese

Pierina Cheung\textsuperscript{a}, David Barner\textsuperscript{b}, and Peggy Li\textsuperscript{c}

\textsuperscript{a}Department of Psychology, University of Waterloo

\textsuperscript{b}University of California, San Diego

\textsuperscript{c}Harvard University

Address Correspondence to:

David Barner
Department of Psychology
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093-0109
p. 858-450-0302
f. 858-734-7190
barner@ucsd.edu
Abstract  When presented with an entity (e.g., a wooden honey-dipper) labeled with a novel noun, how does a listener know that the noun refers to an instance of an object kind (honey-dipper) rather than to a substance kind (wood)? While English speakers draw upon count-mass syntax for clues to the noun’s meaning, linguists have proposed that classifier languages, which lack count-mass syntax, provide other syntactic cues. Three experiments tested Mandarin-speakers’ sensitivity to the diminutive suffix -zi and the general classifier ge when interpreting novel nouns. Experiment 1 found that -zi occurs more frequently with nouns that denote object kinds. Experiment 2 demonstrated Mandarin-speaking adults’ sensitivity to ge and -zi when inferring novel word meanings. Experiment 3 tested Mandarin three- to six-year-olds’ sensitivity to ge. We discuss differences in the developmental course of these cues relative to cues in English, and the impact of this difference to children’s understanding of individuation.

Keywords: individuation, numeral classifiers, Mandarin –zi morpheme, mass nouns, count nouns, bare nouns, word meaning
1. Introduction

When children learn the meanings of words, they observe not only how words relate to things in the world, but also how they are used syntactically. The syntactic status of a word places important constraints on its meaning, thereby limiting the hypothesis space for children learning language. However, the grammatical cues that are available to children vary from one language to another; before syntax can support word learning, children must first learn how it encodes meaning in their language. Here, we investigate the syntactic cues to noun meaning in a classifier language, Mandarin Chinese, by exploring the case study of individuation.

In all attested languages there are words that refer to discrete countable individuals. However, languages vary with respect to how they mark reference to individuals. In many languages, a distinction is made between mass and count syntax. For example, in English, count nouns can occur directly with numerals (e.g., one dog), in singular or plural forms (a dog, some dogs), or with quasi-cardinal determiners (these dogs) that signal reference to sets of individuals. Mass nouns, on the other hand, cannot be used directly with numerals, with singular or plural morphology, or with quasi-cardinal determiners (e.g., *a butter, *three butter).

By many accounts, classifier languages like Chinese, Korean, and Japanese lack a mass-count distinction, and instead only have mass nouns (e.g., Allan 1980; Chierchia 1998). For example, in Mandarin Chinese, nouns cannot co-occur directly with numerals, but require a unitizer (i.e., a “classifier”) for counting, like English mass nouns (e.g. two pieces of toast). Thus, to label three pens in Mandarin requires both the numeral san (three) and the classifier zhi (stick) as in san zhi bi (or “three stick pen”). Also, unlike English count nouns, which obligatorily specify number via singular-plural marking (e.g., a cat vs. some cats), classifier languages normally lack obligatory plural marking. As a result, bare nouns, which lack number marking, are common in
classifier languages.

Here, we do not take a position on whether classifier languages make a mass-count distinction, though such a distinction has been posited both at the level of the classifiers (Cheng and Sybesma 1998, 1999) and the noun (Yi, this issue; see also Chao 1968). Instead, our focus is on how attested grammatical markers of individuation, like classifiers, might be used to learn new words. A large literature on English language acquisition has found that mass-count syntax facilitates the learning of nouns by providing children with cues to individuation. When English-speaking children hear a novel word used in count syntax (e.g., *This is a bicket!*), they assume that it refers to a kind of object rather than an unindividuated substance. This current study examines whether classifiers and other markers of individuation in Mandarin might play a similar role in language learning, and if so, by what age.

In a previous study, Li et al. (2008) examined Mandarin children’s knowledge of how common classifiers relate to individuation. When asked to find “one CL something” (CL = classifier) among several choices, adult participants selected solid, shape-matched objects when presented classifiers like *gen* and *zhang* (e.g., respectively rod shapes or sheet shapes), but rejected things that were non-solid or that didn’t match in shape.¹ However, young children were less successful, and accepted portions of non-solid stuff for these classifiers, so long as the referent matched the shape specified by the classifier (e.g., toothpaste that was shaped like a rod). It was not until approximately six years of age that children began to interpret classifiers like adults, as cues to individuation.

This conclusion, however, is premature, since besides encoding individuation classifiers also

---

¹ Li et al. also tested participants with what Cheng and Sybesma (1999) called “mass classifiers” like *dui* (pile) and *tuan* (wad), which pick out units that are not necessarily discrete physical objects (i.e., portions of stuff of things).
encode item-specific conceptual information, such as the shape, animacy, etc. Thus, many classifiers may emerge late in Mandarin acquisition because of the idiosyncratic information that they encode (Uchida and Imai 1999; Erbaugh 1986). However, not all classifiers encode item-specific meanings. In most classifiers languages, a default classifier also exists, which can be used for novel nouns, or when no specific classifier is available. In Mandarin, for example, the general classifier is *ge*. *Ge* is by far the most frequent classifier in Mandarin, and because it provides no shape information, it may be easier to learn than other classifiers, and thus may serve as an early cue to individuation.

Besides the classifier *ge*, linguists have noted that Mandarin may also have bound morphology that marks individuation, in an item-general way. According to some reports, the diminutive marker –*zi* is a reliable cue to individuation, since it occurs only on nouns that denote object kinds (Doetjes 1997; Sybesma 2007). Others have argued that diminutive markers frequently act as unitizers cross-linguistically, and allow the conversion of mass nouns into count nouns (Wiltschko 2006). Historically, -*zi*, (meaning “son” or “child”) functioned as a diminutive marker. However, simplification of the phonological system during the Han Dynasty, and a movement away from monosyllabic nouns led to the adaptation of -*zi* as an ending for many nouns (Li and Thompson 1981). Although many native speakers are not aware of the function of –*zi*, linguists have observed that many nouns used with -*zi* refer to “concrete, non-abstract things, that can be counted individually” (Dragunov 1960, p. 81; Sybesma 2007). It is possible that speakers of Mandarin are implicitly sensitive to -*zi* as a cue to individuation. As no experiments have ever examined -*zi* and its relation to object individuation, we therefore included -*zi* in our study and compared it to the general classifier *ge*, in the domain of word learning.

The present study investigated whether the general classifier *ge* and the diminutive suffix -*zi*
are cues to individuation in Mandarin. Experiment 1 tested the hypothesis that -zi occurs with nouns that denote individuals by asking whether this was true for commonly used nouns in Mandarin. Experiment 2 then examined how adult speakers interpret both -zi and ge by contrasting them with bare nouns in two word learning tasks. Experiment 3 extended Experiment 2 by testing Mandarin-speaking children's developing comprehension of the classifier ge, in order to establish the role of classifiers in children’s emerging understanding of individuation.

2. Experiment 1

Experiment 1 tested whether the diminutive suffix -zi is more likely to be used with nouns that refer to object kinds than those that refer to substances. To explore this, we used an object-substance judgment task and a zi-rating task.

2.1 Method

2.1.1 Participants

Participants were 52 native Mandarin speakers (M=25.6 years old) recruited from student and staff populations in two universities in Taiwan. Twenty-seven participants performed the object-substance judgment task and the remaining 25 participated in the zi-rating task.

2.1.2 Stimuli and Procedure

A total of 256 frequent nouns in child-directed speech was selected from a list provided by Sandhofer et al. (2000), who culled 50 Mandarin CHILDES transcripts. We excluded proper nouns, pronouns and compound nouns, and -zi was stripped of from 66 nouns that came naturally with it. This list was divided into List A and List B, each with 128 words. Ten words
from each list were added to the other to verify that there were no systematic group differences.  

Object-Substance Judgment. Fourteen participants were assigned to List A and 13 participants to List B. Participants were asked to rate whether each of 138 words referred to an object (wu4ti3), substance (wu4zhi3), both, or neither.

-Zi Rating. Thirteen participants were assigned to List A and 12 participants to List B. Participants were asked to rate how likely -zi could be affixed to each word, using a scale of 1 to 7 (1 being highly acceptable and 7 being highly unacceptable).

2.2 Results

Object-Substance Judgment. Words in which more than 50% of participants provided discordant object-substance judgments were coded as “neither” (36 words, 14.1%). For the remaining words, an average rating was calculated (where 1=object, 2=both, 3=substance). Nouns were categorized as object labels if they received a score lower than 1.5 (149 words, 58.2%), as substance labels if they received a score higher than 2.5 (21 words, 8.2%), and as “both” if they had scores in between (50 words, 19.5%).

-Zi Rating. An average -zi rating, out of 7, was calculated for each word. First, we verified whether the ratings corresponded to how -zi is used in everyday speech. The average rating was 1.5 for the 66 nouns that originally had a -zi ending, suggesting that participants’ ratings were in accord with the usage of -zi.

Second, to test the relation between object-substance judgments and -zi ratings, we compared the two rating tasks. Nouns categorized as object words received the lowest average -zi rating (2.96), whereas those categorized as substance words received the highest (6.00). Participants showed remarkable agreement for the 20 overlapping words: correlations between the two groups of participants were high for both object-substance ratings (Pearson’s r = .95, p < .0001) and -zi rating (Pearson’s r = .96, p < .0001).
found it acceptable to affix –zi to nouns denoting object kinds but not substance kinds. The -zi ratings of the object noun category differed significantly from the ratings of the substance nouns (2.96 vs. 6.00; t(168)=−6.26), of the nouns rated as “both” (2.96 vs. 5.04; t(196)=−6.02), and “neither” (2.96 vs. 5.29; t(183)=−5.96, all p’s<0.001).

To further explore this finding, we compared object ratings based on how likely a word took –zi as an ending. Since words that naturally occur with -zi received a score of 1.5, we used 2 as the cut-off criterion, resulting in 88 nouns that are compatible with -zi. A majority have object kind (N=78) but not substance kind meanings (N=1). Note that having the -zi suffix is not a necessary condition for referring to object kinds. Out of all nouns that were categorized as object words, participants rated 47.7% of them as unlikely to have a -zi suffix.

2.3 Discussion

Experiment 1 revealed two main findings. First, the object-substance judgment task revealed that many Mandarin nouns denote object kinds, consistent with what is typically found for mass-count languages such as English (Barner et al. 2009; Samuelson and Smith 1999). Second, we found that when a word refers to an object, it is more likely to take the -zi suffix, whereas substance words almost never take a -zi ending. These results support the contention that -zi signals individuation in Mandarin Chinese. Nonetheless, as compared to singular-plural morphology in English, which is obligatory for all count nouns, -zi is a linguistic device that only applies to a subset of individuating nouns.

3. Experiment 2

Experiment 1 indicated that -zi is a potential cue to individuation in Mandarin Chinese. However, to verify this, a word learning experiment is required, to determine whether the presence of -zi causes speakers to infer that novel words refer to objects. We investigated this
question in Experiment 2 by testing adult Mandarin speakers. Also, following a similar logic, we asked whether novel words used with ge are more likely to individuate. To test these questions, we used two tasks: a word extension task and a quantity judgment task.

3.1 Method

3.1.1 Participants

Participants were 48 native Mandarin speakers (M=19.7 years old) recruited from the University of Toronto’s Psychology subject pool. Participants were assigned to one of three conditions, with 16 participants each: bare noun, classifier ge, or the diminutive –zi.

3.1.2 Stimuli and Procedure

We used six sets of hand-crafted stimuli, which were simple-shape solid objects designed to be unrecognizable as known artifacts or substances. At the beginning of each trial set, there was a familiarization phase in which the participant was introduced to a standard object that was named four times with a novel term (e.g., fen2yan2 in the bare noun condition; yi2 ge4 fen2yan2 in the classifier condition; fen2yan2 zi in the diminutive condition). They were then randomly presented with either a block of Word Extension or Quantity Judgment trials, which were counterbalanced between subjects.

Word extension. Participants were shown a shape alternative (that matched in shape but not in substance) and a substance alternative (that matched in substance but not in shape). In the bare noun condition, participants were asked to “point to blicket”; in the classifier condition, “point to ___

---

3 Participants also spoke English as a second language. If speaking English leads to a decreased proficiency in Mandarin, this could work against finding a sensitivity effect towards –zi and ge. However, as we shall see, we found an effect nonetheless.
one-CL blicket”; in the diminutive condition, -zi was suffixed at the end of each novel word: “point to blicket-zi”. Following previous studies, we reasoned that judgments based on shape instead of substance indicate that the novel word refers to an object kind.

**Quantity judgment.** Participants were shown two action figures (Farmer Tom and Captain Peter). One figure had two standard items, and the other had four miniature versions of the item with a lesser total volume or mass. In the bare noun condition, they were asked, “Who has more blicket?”; in the classifier condition, “Who has more CL-blicket?”; in the diminutive condition “Who has more blicket-zi?”). Following previous studies, we reasoned that quantification by number indicates that the noun refers to a kind of individual.

3.2 Results

![Chart](chart.png)

**Figure 1.** Mandarin speaking adults’ performance on Word Extension and Quantity Judgment in bare noun, classifier, and diminutive suffix conditions.

For Word Extension, an ANOVA with three between subjects factors found a significant main effect of condition (bare noun vs. ge vs. –zi; \(F(1,36)=135.69, p<0.038\)), but no main effects or interactions involving block order or item order (see Figure 1). There was no difference
between the ge condition and the -zi condition (64.6% vs. 64.6%; \( t(30)=0, p=1.0 \)). However, compared to the bare noun condition (37.5%), adults extended words by shape significantly more in both the ge condition (64.6%; \( t(30)=2.46, p<0.020 \)) and the -zi condition (64.6%; \( t(30)=2.35, p<0.026 \)). Quantity judgment showed similar results. Only a main effect of condition \( (F(1,36)=93.85, p<0.021) \) was found. Adults quantified significantly more by number for ge (60.4%; \( t(30)=2.23, p<0.033 \)) and -zi (71.9%; \( t(30)=2.99, p<0.005 \)) than for bare nouns (29.2%).

3.3 Discussion

For adult speakers of Mandarin, nouns were more likely to individuate when used with the classifier ge and the diminutive suffix -zi, as measured by both word extension and quantity judgment. These results suggest that both ge and -zi offer cues to individuation. In Experiment 3, we explored whether children use ge to infer individuation when learning novel nouns.

4. Experiment 3

In Experiment 3, we tested children’s sensitivity to grammatical markers of individuation, by testing whether the classifier ge affects word learning. Previous studies report that children begin to produce classifiers by 2-and-a-half years of age, and reach adult-like proficiency by the age of 6 or 7 (e.g., Chien et al. 2003; Erbaugh 1986; Li et al. 2008). However, no previous study has tested children’s sensitivity to ge, which is the most frequently used classifier in children’s language, and the first to emerge in their speech.

4.1 Method

4.1.1 Participants

Participants were native speakers of Mandarin, including 31 3-year-olds (M=3;6, range=3;1-3;11), 32 4-year-olds (M=4;6, range=4;1-4;12), 32 5-year-olds (M=5;7, range= 5;0-6;0), and 31 6-year-olds (M=6;7, range=6;1-6;10) recruited from daycares in Taiwan.
4.1.2 Stimuli and Procedure

Children were assigned to one of two conditions (bare noun and classifier), and were tested on the word extension task with the stimuli and procedure of Experiment 2.

4.2 Result

![Graph showing percentage of shape judgments for bare noun and classifier conditions across different age groups.]

**Figure 2.** Percentage of judgments based on shape for the bare noun and classifier conditions in Mandarin-speaking 3-, 4-, 5-, 6-year olds and adults.

An ANOVA of the children’s data with three between-subject variables found a marginal effect of condition (bare noun vs. classifier) on word extension ($F(1,110)=638.7, p<0.058$). We also found marginally significant effects of age group ($p<0.062$) and item order ($p<0.058$). No single age group showed a significant effect of condition (i.e., classifier use), suggesting that children are overall not very sensitive to the general classifier as a marker of individuation. However, the difference in percentage of shape judgments between the bare noun and classifier conditions became gradually bigger from 3-year-olds (3.19%) to 6-year-olds (20.0%; see Figure 2), suggesting that as children become older, they may become more likely to show sensitivity to the classifier *ge* as a cue for individuation.
5. General Discussion

The current study tested two potential syntactic cues to individuation in Mandarin Chinese. Using two noun-judgment tasks, Experiment 1 found that adults were more likely to affix -zi to a noun if it was categorized as an object denoting word, suggesting that –zi is a potential cue for individuation. Experiment 2 tested whether adults were sensitive to ge and -zi when inferring whether novel nouns referred to individuals. Results from word extension and quantity judgment provided strong support for this claim. Experiment 3 found, in contrast, that Mandarin-speaking children were relatively insensitive to ge when learning words, suggesting that classifiers may not be instrumental in learning about individuation early in acquisition.

Together, these results suggest that grammatical morphemes like ge and possibly -zi gradually become associated with individuation in acquisition, but may not contribute critically to the meanings of nouns in acquisition. We recently demonstrated, using the quantity judgement task for known nouns, that Mandarin-speaking children who lack classifier understanding nonetheless learned the appropriate meanings of nouns. For example, when asked whether Farmer Tom and Captain Peter has more with object nouns (e.g., ”Who has more ball?”), children pick by number (i.e., 4 tiny balls is more than 2 gigantic ones). When asked about substance nouns (e.g., ”Who has more sugar?”), children pick by volume (i.e., 2 gigantic piles is more than 4 tiny piles). Clearly, children know which nouns individuate and which do not. Therefore, classifiers and other grammatical markers like -zi may reflect existing meanings, rather than add content to nouns.

Recent studies of noun interpretation in English, Japanese, and Mandarin are consistent with this conclusion, and suggest that nouns across these languages do not differ in their ability to individuate (Barner et al. 2009; Li et al. 2009). Although children learning English are more
likely to think that a novel noun denotes a kind of object than learners of Japanese (Imai and Gentner 1997; Imai and Mazuka 2007; Soja et al. 1990), such effects are likely mediated by syntactic inferences rather than by conceptual or semantic differences between the languages. Since English-speaking children must syntactically parse novel nouns as either mass or count, and since count nouns are more frequent, when they hear a novel noun that is used in ambiguous syntax, they may assume that the word is a count noun (and thus assume that it individuates). However, children learning languages like Japanese or Mandarin make no such syntactic inferences. As we have shown here, although classifier languages like Mandarin provide some cues to individuation, these cues are not available in early acquisition, and thus are not able to affect early word learning.

To conclude, although grammatical markers such as *ge* and *–zi* in Mandarin Chinese become associated with individuation by adulthood, they may not necessarily be a cue to individuation that children use as they initially learn nouns. Linguistic devices such as classifiers and count syntax may simply reflect a deeper conceptual distinction between objects and substances, but do not modify our underlying cognitive representation or create structures that guide how things are perceived.

**Acknowledgments** Special thanks to Lisa Cheng and Rint Sybesma for discussions on *–zi*, Frankie Chen for his help with preparing the word list in Experiment 1, and Lichun Chang, Becky Huang, and Xiaowen Xu for their help with data collection in Taiwan. We are grateful to the parents and children who participated, and to the daycares and preschools in Taipei for their support and help with recruitment. This work was supported by a Connaught Grant awarded to D.B. and NIH NRSA to P.L.
References


Imai, M. and R. Mazuka. 2007. ‘Language-relative construal of individuation constrained by


