An intermediate stage of spec-to-head reanalysis: evidence from Mandarin and Cantonese

Shuyan Wang
University of Connecticut

1. Introduction

Mandarin Yi ‘one’ has been suggested to have grammaticalized from a numeral to an indefinite article (e.g., Chen 2003; Zhang 2013), but no syntactic analysis of the grammaticalization process has been provided. This paper will go beyond previous analyses and systematically explore the grammaticalization process of Mandarin yi. In particular, I will propose that the grammaticalization of Mandarin yi is an ongoing process and yi is ambiguous between a numeral and an indefinite article, and that the grammaticalization of Mandarin yi is an instantiation of a well-known phenomenon: specifier-to-head reanalysis (e.g., Lyons 1999; Wood 2003; van Geldern 2001, 2004). I will further argue that there is an intermediate stage of grammaticalization: a numeral ‘one’ starts as a specifier of a phrase; then, it adjoins to a head before it is reanalyzed as an independent head projecting a separate phrase. In other words, at this intermediate stage, the indefinite article ‘one’ is head-adjoined to another head. Cantonese jat ‘one’ and Turkish bir ‘one’ will provide evidence for this intermediate stage.

The paper is organized as follows. Section 2 will discuss the unique characteristics of Mandarin yi ‘one’. A new account will be proposed in Section 3 to cover the characteristics of Mandarin yi. In particular, I will discuss the spec-to-head reanalysis of Mandarin yi. Then Section 4 will describe Cantonese jat ‘one’ and discuss the intermediate stage of grammaticalization. I will then provide Turkish data to support the proposal of an intermediate stage. Section 5 will conclude this paper.

2. Ambiguity of yi

The possibility of the existence of indefinite articles in Mandarin has been discussed in some works (e.g., Wang 1989; Chen 2003). For instance, Wang (1989) suggests that in Old Chinese, when the classifier ge occurs in Cl-N phrases, it can serve as an indefinite article. In this respect, also relevant is Chen (2003), who discusses a grammaticalization process involving numeral yi ‘one’ and classifiers in Mandarin. Chen (2003) suggests that Mandarin has an indefinite article. In particular, he suggests that in yi-Cl-N phrases, ‘yi+Cl’ (yi ‘one’ and a classifier) together can be used as an indefinite article, while in Cl-N phrases the classifier alone is an indefinite article.

---

1 Note that Chen (2003) simply suggests that ‘yi+Cl’ semantically function as an indefinite article, in that ‘yi+Cl’ can be used in the situations where an indefinite article (e.g., a/an in English) is usually used. No specific syntactic analysis is, however, given in the paper.
Following Chen (2003), I argue that yi has grammaticalized from a numeral to an indefinite article and that yi is ambiguous between a numeral and an indefinite article. The ambiguity analysis can cover a range of observations. First, numeral expressions with yi ‘one’ in Mandarin can be interpreted as indefinite specific or non-specific, while numeral expressions with other numerals are usually interpreted only as indefinite non-specific (e.g., Huang 1987, Tsai 1996). The Numeral (Num)-Classifier (Cl)-Noun (N) phrases, like san-ge xuesheng (three-Cl-student) in (1), are standardly referred to as numeral expressions in Mandarin. Two types of numeral expressions are assumed to exist in Mandarin (e.g., Li 1998): individual-denoting and quantity-denoting expressions. The former refers to some entities/individuals, while the latter expresses the quantity of entities/individuals (e.g., Li 1998). For instance, the numeral expression in (1) refers to some students. The numeral expression in (2) just denotes the quantity and the verb expresses the sufficiency of a certain amount.

(1). wo kandao san-ge xuesheng. 
’I saw three students.’
(2). san-ge xuesheng bu gou. 
’two-Cl student not enough’
’three-Cl student not enough’
‘Three students is not enough.’ (Huang et al. 2009)

Individual-denoting numeral expressions in Mandarin are generally considered to be indefinite non-specific expressions (e.g., Huang et al. 2009). They usually do not appear in subject or topic positions (see (3b) and (3c)), since these positions in Mandarin do not allow non-specific readings (e.g., Chao 1968; Li & Thompson 1981; Lee 1986; Li 1996).

(3). a. wo kandao-le san-ge xuesheng. 
i see-ASP two-Cl student
’I saw two students.’

b. ??san-ge xuesheng chi-le dangao. 
three-Cl student eat-ASP cake
‘Three students ate the cake.’

As a quantity-denoting expression, a numeral phrase can occur in these positions. However, quantity expressions show different structural properties from individual-denoting numeral expressions. For example, quantity-denoting expressions cannot co-refer with a pronoun, as in (4). I will not go into these issues here, since they are not relevant to the main topic of the paper (for relevant discussion, see Li (1998) and Huang et al. (2009)). In this paper, I will focus on individual-denoting numeral expressions. The term ‘numeral expression(s)’ will be used to refer only to individual-denoting numeral expressions.
As mentioned above, numeral expressions in Mandarin are indefinite non-specific expressions and are disallowed in subject/topic positions, as in (3). However, *yi ‘one’*-Cl-N phrases are different in this respect. They can be either specific or non-specific. Thus, Tsai (1996) noted that numeral expressions with *yi ‘one’*, but not other numerals like ‘two’ or ‘three’, can be specific (see also Huang 1987). As a result, *yi*-Cl-N phrases, unlike other numeral expressions, can occur in subject positions, as in (5).

(5). yi-ge xuesheng chi-le dangao.
   one-Cl student eat-ASP cake
   ‘A student ate the cake.’

These data suggest that numeral expressions with *yi ‘one’* in Mandarin can be interpreted as indefinite specific or non-specific, while numeral expressions with other numerals are usually interpreted only as indefinite non-specific. Then the question is why expressions with *yi ‘one’* is different from other numeral expressions in this aspect. In fact, the proposal that *yi* is ambiguous between a numeral and an indefinite article can account for these data. I argue that *yi* in (5) actually is an article, not a numeral. As discussed above, numeral expressions in Mandarin are indefinite non-specific, so they cannot appear in subject/topic positions. However, *yi* can function as an indefinite article. Then it is not surprising that expressions with an indefinite article can be interpreted as indefinite specific and can occur as a subject or a topic in Mandarin.

Another unique feature of Mandarin *yi* is that when *yi*-Cl-N phrases stay under the scope of negation, a ‘not any’ interpretation may emerge. As illustrated by (6), a *yi*-Cl-N phrase stays under the scope of negation (i.e., *congmei ‘never’*) and then they can be interpreted as ‘not any’ (i.e., ‘Xiaohong has never had any boyfriend’).

(6). Xiaohong congmei jiao-guo yi-ge nanpengyou.
   Xiaohong never make-EXP a-Cl boy.friend
   ‘Xiaohong never had any boyfriend.’

I argue that with the ‘not any’ interpretation, *yi* in (6) is an indefinite article, not a numeral. Similar patterns are found in English. As shown by (7a), when the object contains an indefinite article (i.e., *a boyfriend*), the sentence can be interpreted as ‘Mary has never had any boyfriend’. However, if the object is a numeral expression (i.e., *one boyfriend*), the ‘not any’ interpretation disappears, as in (7b).

(7). a. Mary has never had a boyfriend.
   ‘Mary has never had any boyfriend.’
   b. *Mary has never had one boyfriend.
   Intended meaning: ‘Mary has never had any boyfriend.’
In summary, Mandarin yi demonstrates some unique characteristics, and differs from other numerals in some aspects. Numerals expressions in Mandarin are interpreted as indefinite non-specific. However, expressions with yi ‘one’ can be interpreted as specific. Furthermore, yi-Cl-N phrases under the scope of negation be interpreted as ‘not any’. This parallels the phrases with an indefinite article in English. I have argued that these unique features of Mandarin yi can be explained by its ambiguity status. In particular, I proposed that in these two special cases, yi functions as an indefinite article, rather than a numeral. The next section will focus on the grammaticalization process of Mandarin yi and I will propose syntactic analyses for the two different yi.

3. Specifier-to-head reanalysis

In terms of the diachronic reanalysis over the grammaticalization process, I propose that yi ‘one’ as a numeral stays in the specifier of CIP (i.e., Spec CIP), and then the grammaticalization process leads to a reanalysis as an indefinite article, which heads a Quantifier Phrase. As shown in (8a), numerals are located in the Spec CIP (following e.g., Tang (1990), I assume that a classifier heads its own projection above NP), while the indefinite article yi ‘one’ is a head projecting a Quantifier Phrase, as in (8b).

(8). a. 
```
CIP
   Num    Cl'
     Cl   NP
```

b. 
```
QP
   Q'   NP
      Q  yi
```

It should be noted that we may be dealing here with a more general grammaticalization process. Thus, by examining diachronic variations between Old English, Middle English, and Modern English, van Geldern (2001, 2004) argues that many functional categories developed as a result of a change of a specifier of a phrase to a head that projects a separate phrase. In fact, similar claims have also been made for many elements within CP (like complementizers) and DP, like demonstratives (see also Lyons 1999; Wood 2003; van Geldern 2001; Huddleston & Pullum 2002; Willis 2007, among others).

As discussed above, I argue that yi ‘one’ in Mandarin is ambiguous between an indefinite article and a numeral. As shown by (8a), a numeral is in the specifier of CIP, so its existence is closely related to the presence of a classifier. In other words, whenever there is a numeral, a
Classifier Phrase is projected; a numeral cannot occur without a Classifier Phrase. In contrast, as displayed in (8b), the indefinite article (i.e., yi ‘one’ in the indefinite article usage) is the head of QP, whose existence is not dependent on the presence of a classifier. This predicts that expressions where indefinite article yi ‘one’ (not the numeral yi ‘one’) is directly followed by a noun may be possible in Mandarin. Because there is no numeral, there is no need for a Classifier Phrase. This is borne out. H. Huang (1981) and Lü et al. (1999)\(^2\) observed that yi can directly combine with bare nouns without a classifier, while other numerals cannot, as in (9).

(9). a. yi nanhai  
   one boy
   ‘a boy’

   b. san*(ge) nanha  
   three-Cl boy
   ‘three boys’

Furthermore, numeral phrases in Mandarin, but not yi-N phrases, can be used to answer ‘how many’ questions. As illustrated by (10a) and (10b), numeral expressions (i.e., yi-ben shu and san-ben shu) can be used as an answer to a ‘how many’ question. However, yi-N phrases cannot be used as answers here, as in (10c).

(10). ni mai-le duoshao shu?  
      you buy-ASP how many book
      ‘How many books did you buy?’

   a. wo mai-le yi-ben shu.  
      I buy-ASP one-Cl book
      ‘I bought one book.’

   b. wo mai-le san-ben shu.  
      I buy-ASP three-Cl book
      ‘I bought three books.’

   c. *wo mai-le yi shu.  
      I buy-ASP one book
      ‘I bought a book.’

Similar patterns are found in English: numerals can be used to answer ‘how many’ questions, while indefinite articles (i.e., a/an) are degraded as an answer here. As shown in (11), numerals, like one and three, can be used to answer the ‘how many’ question (see (11a) and (11b)), while the indefinite article cannot be used here, as in (11c). Thus, the impossibility of Mandarin yi-N phrases as an answer to ‘how many’ questions provides further support for the current analysis in which yi is an indefinite article in yi-N phrases.

(11). How many books did you buy?
      a. I bought one book.
      b. I bought three books.

\(^2\) H. Huang (1981) and Lü et al. (1999) do not provide an analysis for yi-N phrases. They only briefly mentions that when the numeral is yi ‘one’, the classifier sometimes can be optional.
In summary, this section has investigated the grammaticalization process. In particular, I argued that Mandarin yi ‘one’, as a numeral, is located in the specifier of a Classifier Phrase; a grammaticalization process leads to its reanalysis as the head of a Quantifier Phrase. This diachronic reanalysis is an instantiation of a well-known phenomenon: specifier-to-head reanalysis (e.g., Lyons 1999; Wood 2003; van Geldern 2001, 2004). The following section will propose an intermediate stage of grammaticalization with evidence from Cantonese.

4. **Intermediate stage of grammaticalization**

In this section, I will argue that jat ‘one’ in Cantonese is also undergoing the grammaticalization process from a numeral to an indefinite article. However, I will further argue that jat has not grammaticalized as much as yi in Mandarin: jat has not been reanalyzed as an independent head projecting a separate phrase; instead, jat is head-adjoined to the classifier head.

Numeral expressions in Cantonese are usually interpreted only as indefinite non-specific, while numerals expressions with jat ‘one’ can be interpreted as indefinite specific or non-specific. As illustrated by (12a), numeral expressions in Cantonese do not appear in subject/topic positions, since these positions usually do not allow indefinite non-specific expressions. However, a jat-Cl-N phrase in the subject/topic position is much better, as shown in (12b).

(12). a. ??saam-go hoksaang mei lei. b. jat-go hoksaang mei lei.
   three-Cl student not come a-Cl student not come
   ‘Three students didn’t show up.’ ‘A student didn’t show up.’

I argue that jat in (12b) actually is an article, not a numeral. Numerals expressions in Cantonese are indefinite non-specific, so they cannot appear in subject/topic positions. If jat can function as an indefinite article, then it is not surprising that jat-Cl-N phrases differ from other numeral expressions.

Furthermore, when jat-Cl-N phrases stay under the scope of negation, the ‘not any’ interpretation emerges, as in (13). As discussed above, Mandarin yi and English a/an show similar patterns, as in (6) and (7). This provides more support for the proposal that jat in (13) is an indefinite article, not a numeral.

(13). Siuhung chung-mut gaau jat-go naampangyau.
   Siuhung never made a-Cl boy.friend
   ‘Siuhung has never had any boyfriend.’

Therefore, Cantonese jat ‘one’ differs from other numerals in various aspects. Based on its unique features, I argue that jat in Cantonese is ambiguous between a numeral and an indefinite article. In these special cases (e.g., in subject/topic positions; interpreted as ‘not any’), jat function as an indefinite article, rather than a numeral. Then the next question is what the syntactic structure of jat is. If we follow the proposal for Mandarin, the structure should be like (14). In this structure,
numerals stay in Spec ClP, but the indefinite article *jat ‘one’* has been diachronically reanalyzed as the head of a QP, as in (14b).

(14). a. \[\text{ClP samm } [\text{Cl' -go } [\text{NP hoksaang}]] \text{ three Cl student} \]
   \[\text{‘three students’} \]

   b. \[\text{QP } [\text{Q' jat } [\text{ClP } [\text{Cl' -go } [\text{NP hoksaang}]]] \text{ a Cl student} \]
   \[\text{‘a student’} \]

According to the structures proposed in (14), the article *jat* then would not be dependent on the presence of a classifier and should be directly combinable with a noun. As discussed above, Mandarin *yi* can combine with a noun without a classifier in between. However, it is not the same case for Cantonese *jat*. Cantonese *jat* always needs a classifier, as in (15).

(15). \[\text{jat-*(go) hoksaang} \]
   \[\text{a/one-Cl student} \]
   \[\text{‘a student/one student’} \]

One potential explanation for this cross-linguistic variation is that *jat* in Cantonese and *yi* in Mandarin have grammaticalized to different degrees. *Jat ‘one’* may not have grammaticalized as much as *yi ‘one’*. Then, even when *jat ‘one’* functions as an indefinite article, it still requires the presence of a classifier. Based on this, I argue that the current structure for Cantonese numeral phrases should be like (16). In (16a), numerals, including the numeral *jat ‘one’*, stay in Spec ClP. On the other hand, the article *jat* is adjoined to the classifier head, as in (16b). The expectation is that in a further stage of grammaticalization, the indefinite article *jat* will be reanalyzed as a head heading its own projection, as in Mandarin. At that time, *jat-N* phrases may be expected to emerge in Cantonese.

(16). a. \[
    \text{CIP} \\
    \text{Num} \\
    \text{Cl' } \\
    \text{Cl NP numerals} \\
\]

   b. \[
    \text{CIP} \\
    \text{Cl' } \\
    \text{Cl NP } \text{jat-Cl} \\
\]

As discussed above, van Geldern (2001, 2004) shows that many functional categories developed from a specifier of a phrase to a head that projects a separate phrase (see also Lyons 1999; Wood 2003; Huddleston & Pullum 2002; Willis 2007). Here I propose that before being reanalyzed as a head that can project by itself, a functional category first adjoins to another head. Applying this more general pattern of grammaticalization, numeral ‘one’ starts as a specifier of a phrase; then, in the grammaticalization process, it adjoins to a head before it is reanalyzed as a head that can project a phrase by itself. Regarding why this intermediate stage exists, it is standardly assumed that when a specifier is merged into the structure, its sister projects. I suggest that when a specifier
is reanalyzed as a head, initially it cannot project, which means at this point it cannot take a complement. As a result, the relevant element is first adjoined to another head.

Turkish provides more evidence for the proposed intermediate stage. Turkish has been assumed to have an indefinite article which has grammaticalized from the numeral ‘one’. It is worthwhile exploring whether Turkish, Mandarin, and Cantonese share any common features in the relevant respects.

Turkish has a unique way to distinguish the indefinite article bir from the numeral bir: the ordering of adjectives and bir. When bir immediately precedes the nominal head, it is interpreted as an indefinite article, as in (17a). On the other hand, when bir is separated from the head noun by adjectives, it is interpreted as a numeral rather than an article, as in (17b).

(17). a. iyi yeni bir kitap
   good new a book
   ‘a good new book’
   ‘*one good new book’
   ‘*a good new book’ (Yukseker 2000)

The ordering of adjectives and bir shown in (17) suggests that the indefinite article bir cannot be separated from the head noun. Therefore, I propose that Turkish numerals stay at Spec of Numeral Phrase (18a), while the article bir is head-adjoined to Noun head (18b). As a result, the indefinite article bir cannot be separated from the head noun by any other elements.

(18). a. [NumP bir [NumP yeni [NP kitap] one new book]
       [NP yeni [NP bir kitap] a book]

The fact that the indefinite article bir in Turkish cannot be separated from the head noun can be taken as providing supporting evidence for the ‘head-adjunction’ analysis. Furthermore, based on the proposed structures for Mandarin yi, Cantonese jat, and Turkish bir, I argue that over the grammaticalization process from a numeral ‘one’ to an indefinite article, ‘one’ needs to first adjoin to a head before it can be reanalyzed as a head heading an independent Quantifier Phrase. It is then expected that in a further stage of grammaticalization, both Cantonese jat ‘one’ and Turkish bir ‘one’ will be reanalyzed as a head heading a Quantifier Phrase. At that time, Turkish bir, as a true indefinite article, should be able to be separated from the head noun by other elements (like adjectives), because it will no longer be head-adjointed to the Noun head.

---

3 It is actually not completely clear if yi in Mandarin, as an indefinite article, is adjoined to the Noun head or not. As shown in (i), adjectives occurring between yi-Cl and a noun can appear with de or without de.

i. yi-suo meiguo (de) gaoxiao
   one-Cl American DE university
   ‘a/one American university’

However, there is speaker variation regarding whether de can be present in yi-N phrases. As shown in (ii), adjectives without de are uniformly acceptable in yi-N phrases. But when de is present here, there is speaker variation (indicated by #).

ii. yi meiguo (#de) gaoxiao

Bošković and Hsieh (2013) suggest that adjectives with de are adjoined to NP, while adjectives without de are head-adjointed to the Noun head. If this is correct, it is possible that for the speakers who allow examples like (ii), yi as an
5. Conclusion

This paper has systematically discussed the grammaticalization process of Mandarin yi. In particular, I proposed that Mandarin yi is ambiguous between a numeral and an indefinite article, and that the grammaticalization of Mandarin yi is an instantiation of a well-known phenomenon: specifier-to-head reanalysis (e.g., Lyons 1999; Wood 2003; van Geldern 2001, 2004). Then I argued that Cantonese jat ‘one’ is also undergoing a grammaticalization process from a numeral to an indefinite article. However, Cantonese jat has not grammaticalized as much as Mandarin yi. In particular, I argued that Cantonese jat is at an intermediate stage of grammaticalization: a numeral ‘one’ starts as a specifier of a phrase; then, it adjoins to a head before it is reanalyzed as an independent head projecting a separate phrase. Turkish bir ‘one provides more evidence for this intermediate stage.

References
