# *De* 的 as an underspecified classifier: first explorations

Lisa Lai-Shen Cheng and Rint Sybesma
UNIVERSITEIT LEIDEN

*for Noam Chomsky*


## 1. Introduction: what are the issues?

For several reasons (to be made precise below), it has proven very hard to pin down the exact status and function of the element *de*, which we find in the nominal phrase in Mandarin between the noun and its modifier (which precedes the noun). As a result, it has been called by many different names (such as "marker of explicit modification" (Chao 1968, 285), and "associative" marker (Li and Thompson 1981, 111) and it has been analysed in many different ways: as a $C^0$ or complementizer (Cheng 1986), as a kind of clause-type marker (Cheng 1998), as a $D^0$ (Simpson 2002, à la Kayne 1994), as the head of a ModP (Rubin 2003), as a marker of predicate inversion having taken place (Den Dikken and Singhapreecha 2004), as a marker of the division of the NP in two syntactico-semantic domains (Paul 2005), as a semantic type-lowerer (Huang 2006), as an element that signals the presence of a general abstracton operator (Cheng and Sybesma 2006), or as any combination of any of these — to mention just some of the more influential or relatively recent proposals. Very recently, Arsenijevic and Sio (2007, 2008) have argued that *ge3*, the Cantonese counterpart of *de*, is a type of classifier. Our proposal, though different in the details, is in the same vein.

There is a lot of literature on *de*, especially in Chinese; see the bibliography in Xu (2006); in particular, the series of papers by Zhu Dexi, starting in the late fifties, has been quite influential. Many of Zhu's observations regarding the distribution of *de* and how its absence or presence affects the semantics of the phrase are to be found in almost every paper on the subject, including the present one. Other important sources on the subject include Paris (1979).

One reason why it is hard to determine what *de* is, is that we find it with every possible type of modifier: simple adjectives, so-called complex adjectives (terminology due to Zhu Dexi; see below), possessors, different types of relative clauses, prepositional phrases, and several (other) non-predicative modifiers; examples are given in (1).

(1)    a.    dà de yú                                 simple adjective
                  big DE fish
                  'big fish'
        b.    fēicháng      dà de yú                  complex adjective
                  extraordinarily big DE fish
                  'very big fish'
        c.    Zhāng Sān de yīfu                   possessor
                  Zhang San DE clothes
                  'Zhang San's clothes'
        d.    méi mǎi-guo shū de rén            R(elative) C(lause)

NEG buy-EXP  book DE person
'people who have never bought a book'

e.  tā chàng gē    de shēngyīn                    gapless RC
    3S sing   song DE voice
    'the voice with which he sings'

f.  duì      érzi de tàidu                        PP
    regarding son DE attitude
    'the attitude towards his son'

g.  yǐqián  de zǒngtǒng                           non-predicative modifier
    former DE president
    'the former president'

What complicates matters is that [modifier + *de*] does not exclusively precede a bare noun. We find it preceding phrases containing a numeral or demonstrative as well:

(2)  a.  Zhāng Sān de  yī-jiàn yīfu
         Zhang San DE one-CL clothes
         'one piece of clothing of Zhang San's'

     b.  Zhāng Sān de  nèi-jiàn yīfu
         Zhang San DE that-CL  clothes
         'that piece of clothing of Zhang San's'

Some of the discussion in the literature is about differences and similarities between cases in (2) and their respective counterparts in (3); the consensus seems to be that there is a contrastive focus reading in (2) (regarding *Zhāng Sān*), which is absent in (3).

(3)  a.  yī-jiàn  Zhāng Sān de yīfu
         one-CL Zhang San DE clothes
         'one piece of clothing of Zhang San's'

     b.  nèi-jiàn Zhāng Sān de yīfu
         that-CL  Zhang San DE clothes
         'that piece of clothing of Zhang San's'

Another complicating factor is that *de* is optional in some cases but not in others. There are cases which are easy to define, such as in the case of kinship terms modified by a pronoun (as illustrated in (4)).[1] Note that even in these cases, we have to take into account that *de* is optional only in the sense that its presence or absence does not affect the grammaticality and not in the sense that it makes absolutely no difference to the meaning, however subtle the difference may be (at least according to some researchers; Lu Jianming p.c.); in other words, we may be dealing with different underlying structures. In what follows, we will not focus on these cases.

(4)  a.  tā (de) bàba / Zhāng Sān *(de) bàba
         3s DE   dad / Zhang San   DE   dad

---

[1] In certain cases of stacked modifiers, *de* is optional, except the one closest to the noun, which tends to be obligatory.

'her dad' / 'Zhang San's dad'
b.     tā (de) nán-péngyou  / Zhāng Sān *(de) nán-péngyou
       3s  DE  boy-friend /     Zhang San    DE  boy-friend
       'her boyfriend' / 'Zhang San's boyfriend'

We *will* focus on the pattern of distribution we observe with common nominals modified by different types of modifiers, that is, the general pattern. First, there are the clearcut cases in which *de* is obligatory and the cases in which its presence leads to ungrammaticality, as we will see below. Secondly, there are the cases in which *de* is optional (in the sense just mentioned). It is optional, for instance, in phrases such as (1a) and it is *not* optional in any of the other cases in (1). Similarly, under certain conditions, to be investigated in detail below, *de* is optional in phrases involving modifiers preceding the demonstrative, as in (2b). The question is, why is *de* optional in some cases, and not in others? Investigating its distribution and circumstances under which it is optional hopefully leads to insights into the function and status of *de*.

These distributional issues have been the focus of many papers, including some recent ones. Huang (2006) constitutes an original attempt to explain why (5a) is okay without *de*, while (5b) is only okay with.

(5)     a.     zāng shuǐ
               dirty water
               'dirty water'
        b.     hěn  zāng *(de) shuǐ
               very dirty    DE water
               'very dirty water'

Following the work of Zhu Dexi (1956, 1961 and later), Huang distinguishes between simplex adjectives (SAs) and complex adjectives (CAs). *Dà* 'big' in (1a) and *zāng* 'dirty' in (5a) count as SA, *fēicháng dà* 'extremely big' and *hěn zāng* 'very dirty' in (1b) and (5b) respectively as CA. SAs and CAs, she notices, are in complementary distribution. SAs can modify bare nouns without a marker but they cannot function as predicates without additional material, and CAs are the other way around: they need a marker if employed attributively, but when they are used as predicates, no extra material is needed. Huang concludes that SAs are <e> type elements and CAs are <e,t>. That is why SAs need a type-lifter (from <e> to <e,t>) to be able to function as a predicate, and, since, according to Huang (following Chierchia 1998), bare nouns in Chinese are <e> and in modification structures, types must match, CAs need a type-lowerer (from <e,t> to <e>) to be able to function as a modifier of a bare noun. *De* is the type-lowerer. This way, Huang explains the situation we find in (5) (as well as a number of other facts) in a very original way. There are, however, a number of questions which remain unanswered; we will mention one or two here. A point that Huang does not address is the fact that insertion of *de* in (5a) is legitimate, as illustrated in (6b):

(6)     a.     zāng shuǐ  (= (5a))    dà  yú        cōngmíng rén
               dirty water            big fish      intelligent person
               'dirty water'          'big fish'    'intelligent person'

| | b. | zāng de shuǐ | dà de yú | cōngmíng de rén |
|---|---|---|---|---|
| | | dirty DE water | big DE fish | intelligent DE person |
| | | 'dirty water' | 'big fish' | 'intelligent person' |

If *zāng* and *shuǐ* are both of the same type, we do not expect *de* to show up. The same holds for possession cases such as illustrated in (1c). Presumably, a proper name is also of type <e>, thus nullifying the need for type-shifter *de*. But, in fact, *de* is obligatory in these cases. Finally, the optionality of the cases in (7) is hard to explain for Huang. The examples in (7a) and (7b) are the same (*de* is optional in both cases), but Huang would expect them to be different: (7a) should be fine without *de* as there is no need for a type-lowerer, and in (7b) we do need *de* according to Huang, but we are perfectly happy without.

(7) a. Zhāng Sān (de) nèi-jiàn yīfu
Zhang San DE that-CL clothes
'that piece of clothing of Zhang San's'
b. hěn zāng (de) nèi-jiàn yīfu
very dirty DE that-CL clothes
'that very dirty piece of clothing'

Paul (2005) zooms in on the differences between (6a) and (6b). Summarizing some of the previous literature (aside from works by Zhu Dexi, also Fu 1987, Feng 2001 and others), she concludes that modifiers without *de* have a "defining property" reading and those with *de* have an "accessory property" reading (p. 770). The phrase *cōngmíng rén* 'intelligent person' in (6a), for instance, refers to a kind of person for whom being intelligent is a defining property, while the counterpart with *de*, *cōngmíng de rén* 'intelligent DE person', is a person who happens to also be intelligent. Paul then proposes that the nominal domain is divided into two domains, one below *de* and one above. Modifiers inserted in the lower domain (incidentally, only heads are allowed to be inserted here) give one the defining property reading, ones inserted in the higher domain yield the accessory property reading.

Although Pauls paper presents an interesting new take on these very relevant issues, it leaves us in the dark with respect to the syntactic status of *de*. What is also not clear is how modifiers with and without *de* can appear on top of the demonstrative, as we saw in (7) (in which the modifiers without *de* are not restricted to elements with head-status).

Below we will return to some of the data, problems and insights in papers by Huang and Paul.

In sum, it has become clear that the general issue we are interested in, "the syntax and semantics of *de*", can be seen as consisting of the following sub-issues:

**I.** what is the distribution of modifiers with and those without *de* (and what are the meaning differences, if there is any)?
**II.** what are the structural positions of modifiers with *de* and what are the structural positions of those without?
**IIIa.** what is the internal make-up of modifiers with and those without *de*?

which includes the separate question:
**IIIb.** what is the nature of *de*? and
**IV.** how do these issues interrelate?

In this paper we will not be able to go into all sub-issues equally deeply, but we will discuss them all to a certain extent. With respect to IIIb we will, as mentioned, investigate the hypothesis that *de* (along with its Cantonese counterpart *ge3*) is a type of classifier.


## 2. First step: noun ellipsis

For our investigations, we take as our point of departure the observation that in Mandarin, noun phrase ellipsis (henceforth, NPE) is licensed in two different environments: nouns can be elided immediately following the classifier as well as immediately following *de*. The following examples illustrate this.

(8) a. tā gāngcái chī-le yī-ge píngguǒ, nǐ yě yīnggāi chī yī-ge
    3s just-now eat-PERF one-CL apple, 2s also ought eat one-CL
    'he just ate an apple, you should also eat one'
  b. tā bù xǐhuān nèi-běn shū, tā xǐhuān zhèi-běn
    3s NEG like that-CL book, 3s like this-CL
    'he does not like that book, he likes this one'

(9) a. wǒ xǐhuān hóng-sè de xié, tā xǐhuān huáng-sè de
    1s like red-color DE shoe, 3s like yellow-color DE
    'I like red shoes, he likes yellow ones'
  b. tā zuótiān mǎi-le yī-jiàn xīn de máoyī, wǒ mǎi-le yī-jiàn jiù de
    3s yesterday buy-PERF one-CL new DE sweater, 1s buy-PERF one-CL old DE
    'he bought a new sweater yesterday, I bought an old one'

This has been observed for NPE in Cantonese by Arsenijevic and Sio (2008) as well: NPE can occur after the classifier and after *ge3*.
  This raises the question what classifiers and *de* have in common such that they can both license NPE. One possible answer that comes to mind is that they are both elements of the same category, more precisely, that *de* is also a classifier of some sort.
  In a recent paper, Alexiadou and Gengel (2008) argue that NPE in Romance and Germanic languages such as Italian, Spanish, English, Dutch and German is licensed by a ClassifierPhrase (ClP). The postulation of a classifier phrase in these languages is not new; it goes back at least to Sharvy (1978) and more recent proposals include Borer (2005) and Picallo (2006). Whereas Borer focuses on the quantifying aspects of classifiers (dividing and individuation), Picallo takes the fact that gender represents a form of classification (cf. Croft 1994) as strong support for the idea that Romance and Germanic languages have a ClP in the functional domain of the noun. Alexiadou and Gengel (2008), aside from pointing out that there is an intimate relation between NPE and quantification, more particularly, partitivity, notice that in some Romance and

Germanic languages NPE is licensed by elements which are clearly recognizable as related to gender. Here is an example from Italian (taken from Alexiadou and Gengel):

(10)  a.      un/*uno libro grande è sulla   tavola
              a          book big     is on.the table
              'a big book is on the table'
      b.      uno/*un grande è sulla   tavola
              a          big     is on.the table
              'a big one is on the table'

These examples show that masculin *-o* is obligatory in NPE contexts, and banned elsewhere. Alexiadou and Gengel interpret this as a visible reflex of the ClP, which licenses the ellipsis. They present similar facts from German and Dutch. In English, *one* is (sometimes obligatorily) present in (count noun) NPE and Borer (2005) has made the claim that *one* is one of the elements in English that play a role in individuating. In other words, in all these languages, we see elements in NPE contexts which can easily be associated with the two basic aspects which are generally ascribed to classifiers: classification and individuation. According to Alexiadou and Gengel (2008), then, it is advantageous to assume that all NPE is licensed by a ClP.

That the classifier can license NPE in Mandarin is clear from the facts in (8a,b). The question is where the classifier is in (9a,b). In what follows we will explore the hypothesis that *de* is a classifier.

Arsenijevic and Sio (2007, 2008) have been exploring this hypothesis for Cantonese *ge3*. Arsenijevic and Sio take it that *ge3* is composed of two different components: a deictic demonstrative component and a classifier/division component. The first component, they claim, is directly visible: it is represented by the onset *g*, which has been claimed to be deictic by Fung (2000); the same *g* is part of the distal demonstrative *go2*, the marker for manner adverbial modification *gam2* 'to such a degree', as well as a number of sentence final particles with strong deictic properties. The classifier/division component is deduced exactly on the basis of NPE sentences such as (9a,b). Here are some Cantonese examples (adjusted from examples in Arsenijevic and Sio's handouts):

(11)  a.      nei5 ge3 syu1 hou2-tai2 gwo3 ngo5 ge3
              2SG  GE  book good-read pass 1SG   GE
              'your books are more interesting than mine'
      b.      nei5 bun2 syu1 hou2-tai2 gwo3 ngo5 bun2
              2SG  CL    book good-read pass  1SG  CL
              'your book is more interesting than mine'

Putting aside the interesting question whether there is a deictic component in *de* and *ge3* (if *ge3* has it, *de* must also have it, as these elements behave in virtually the same way), we concentrate on the other component, the classifier part. We will do so in several steps. First, we will clarify what exactly we mean when we say that *de* is a classifier; this is necessary, because it differs from the elements generally classified as classifiers in never immediately following a numeral:

6

(12)    *sān   de rén
         three DE person

We will then investigate a number of facts which seem explainable, once we have acknowledged that *de* is a type of classifier — and *ge3* as well, as we will not limit ourselves to Mandarin. Finally, we will return to the questions in I-IV above.

  Before we do all this, however, we present some data from Thai, as circumstantial evidence. In this language, we find situations in which the modification marker *thîi* (as such the counterpart of *de* and *ge3*) can be replaced by the classifier (data from Den Dikken and Singhapreecha 2004, (32), (33); there is a slight difference in meaning between the two phrases, as Den Dikken and Singhapreecha explain; see also below):

(13)    a.  rôm  thîi jàj săam khan nán
                    umbrella THÎI big three CL that
         b.  rôm  khan jàj săam khan nán
                    umbrella CL big three CL that
                    BOTH: 'those three big umbrellas'

We will discuss these facts more elaborately shortly.


## 3. Classifiers

In the Thai example in (13b) we see two classifiers. They perform different functions. On the basis of the different distributional patterns displayed by the classifier in Mandarin and Cantonese, Sybesma (2007, 2008) argues that classifiers may perform two different functions. To limit the discussion here to the core cases (but see footnote 6), one function is to mark count nouns as count, which Sybesma (2008) calls the "u-marking function" (where "u" is short for "unit"). Note that the u-marking function does not *produce* count nouns; it does not turn a mass noun into a count noun. U-marking is a morpho-syntactic process, not of adding meaning, but of bringing out overtly an aspect of meaning that is already present in the semantic denotation of the noun (or which is given contextually). In some ways it can be seen as a doubling or, maybe, even an agreement operation.[2] In Cantonese the classifier often performs this function. In the Thai example in (13b), as we will discuss further below, the first instance of the classifier is a u-marker.

  The other function classifiers may perform is that of facilitating counting by acting as the go-between between numeral and noun; this is referred to as the "c-marking function" ("c" short for "counting") in Sybesma (2008). Doetjes (1997) argues that numerals and nouns cannot be combined without the intervention of another element.[3] Sybesma (2008) adds that this is only the case for languages in which the noun is not specified for number. Explaining the complementary distribution displayed by classifiers and number morphology, Doetjes shows that the go-between function is in some languages performed by the plural marker, whereas in others, it is taken care of by the

---

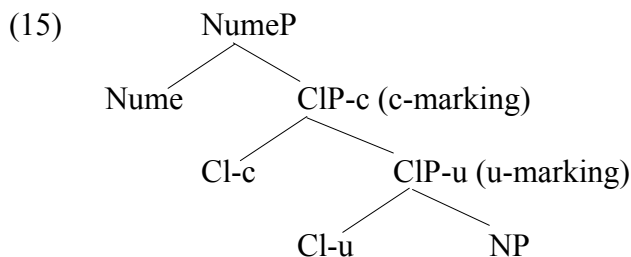[2] Pirani (2007) calls these operations "vacuous derivations".
[3] For Doetjes, this is because the countability of the noun has to be made syntactically accessible to the numeral.

classifier. Mandarin exemplifies the latter case.[4] In the Thai example in (13b), the second instance of the classifier is a c-marker.

A c-marker only shows up in a context of counting: when there is no (overt or covert) numeral, there never is a c-marker. We see this illustrated in Mandarin, where the classifier never occurs without a (sometimes covert) numeral. A u-marker is active in other contexts as well; it is active basically whenever reference to an individual is at play. As a consequence, we see the classifier obligatorily used in definites and specific indefinites in Cantonese, as shown in (14a) below. For reasons we cannot go into now (but see Sybesma 2007, 2008), Mandarin count nouns do not always need a separate u-marker in syntax the way Cantonese nouns do; they have been u-marked in the lexicon (some of them overtly, by the suffix *zi*).

(14) a.     bun2 syu1         (Cantonese)
           CL    book
           'the book'

b.     saam1 bun2 syu1     (Cantonese)
           three   CL   book
           'three books'

c.     sān   běn shū        (Mandarin)
           three CL   book
           'three books'

This means that, in principle, we should be able to sometimes see both the u-marker and the c-marker overtly realized, as in counting, both functions need to be performed: we need the individual having been brought out, and we need the intermediary between the numeral and the count noun. This translates into a structure which may contain, among other layers, these two layers: the layer, right on top of the lexical N, which explicitly marks a noun as countable (u-marking); and the layer which is immediately dominated by the numeral phrase and which is there solely to facilitate counting (c-marking). Let's call both layers ClPs, ClP-c and ClP-u.

(15)          NumeP

     Nume       ClP-c (c-marking)

        Cl-c       ClP-u (u-marking)

            Cl-u         NP

In Chinese languages we don't seem to have the situation in which both slots are filled by the same classifier. The reason for this is the following. Above we have said (following Doetjes 1997) that we need the c-marker as a go-between between the numeral and the noun. Note that although Mandarin nouns need not be u-marked in the syntax because it has already been done in the lexicon, they are unspecified for number. As mentioned

---

[4] For details regarding the relation between these ideas and the work of Borer and Doetjes, see Sybesma (2007, 2008).

above, Sybesma (2008) suggests that numerals can only combine with nouns that are specified for number. As a result, numerals cannot be directly combined with a noun in Mandarin: we need the c-marker, in order to count. Cantonese nouns, in contrast, need the u-marker in syntax, but once we have added the classifier, performing this u-marking function, we have also specified the noun in question for number, since classifiers in general are singular; therefore, we no longer need the c-marker when we count.[5] This way, we only see maximally one marker in both languages. Space limitations prevent us from discussing the question what happens to the slot (in (15)) that is not overtly filled, the u-marker slot (Cl-u) in Mandarin, the c-marker slot (Cl-c) in Cantonese.

      While in Chinese languages we never see the same classifier twice in one noun phrase, in Thai we already saw that we sometimes do. In reference to (13b), we already identified which one is taking care of which function: the one associated with the numeral in both examples in (13) is the c-marker; the other one in (13b) is the u-marker. That the classifier not associated with the numeral is related to the marking of the individual count-units can be seen from the following facts ((16) and (17) based on Kookiattikoon 2001, reproduced in his spelling; thanks to Pornsiri Singhapreecha for explanation and discussion). First there is the contrast in (16):

(16)    a.      som    wan
                  orange sweet
                  'sweet oranges' (as a type of orange)
      b.      som   luk yai
                  orange CL big
                  'big orange(s): oranges that are also/happen to be big'

In (16a) we modify the orange with no reference to the unit in which oranges exist; this is a kind of kind reading. Reference to this unit is not relevant since 'sweet' is a quality of the orange independent of its existence as a discrete entity.

      The following example is a slightly more involved, but points in the same direction. What we see here is reminiscent of Paul's (2005) discussion of the Mandarin facts in (6).

(17)    a.      sat     yai
                  animal big
                  'big animals' (type of animal: elephants, buffaloes, rhinoceroses, etc.)
      b.      sat    tua yai
                  animal CL big
                  'animals that happen to be big' (e.g., a dog that is big for a dog)

In (17a), we are referring to animals that are big as a "defining property", while in (17b) their size is an "accessory property". Of course, when we talk about animals that are big, even as a defining property, we are referring to the units in which they exist. But the difference here is, that in (17a), the size has already been wired in before we start singling out the individuals or units in which they exist. In (17b), we single out the individual first,

---

[5] Cheng (2009) argues that *di1* in Cantonese is a plural classifier, cf. Iljic (1994), who has reasons to doubt the claim that the Mandarin counterpart *xiē* is a plural classifier.

and then say that this particular individual is particularly big. We can have two Danes, but we can also have two big Chiwawas: the two Danes leave out the first *tua* (they are big as a kind), the Chiwawas keep it in (the bigness only applies to certain individuals).

(18)     maa    (tua)   yai     song    tua
         dog    CL-u    big     two     CL-c
         'two big dogs'

In the Chiwawa case, the classifier can be replaced by the marker *thîi*, as in (13):

(19)     maa    thîi    yai     song    tua
         dog    THÎI    big     two     CL-c
         'two big dogs'

What is significant is that it is the u-marker, not the c-marker, which in Thai is interchangeable with the modification marker.

    We would like to propose that the modification markers *de* and *ge3* in Mandarin and Cantonese respectively are classifiers, but the only function they ever perform is the u-marking function of bringing out the unithood. This is all it does; its presence is enforced for purely grammatical-functional reasons. In the following Mandarin and Cantonese phrases, then, the classifier performs the c-marking function and *de* and *ge3* do the u-marking.

(20)   a.   liǎng  běn    hěn    hǎo-kàn         de      shū
            two    CL-c   very   readable        CL-u    book
       b.   loeng5 bun2   hou2   hou2-tai2       ge3     syu1
            two    CL-c   very   readable        CL-u    book
            'two very good books'

Note that *ge3*, as a u-marker, differs from the classifier when it performs the u-marking function in Cantonese: the latter, but not the former, marks the noun for number. As a result, with *ge3*, the noun is unspecified for number, and we need a c-marker when we count. In other words, *ge3* differs from typical classifiers in not having number.[6]

    If this is right, we predict that we cannot have two syntactic u-markers in one and the same phrase. In other words, we rightly predict the sentence in (21a) to be

---

[6] In phrases involving the modification of a mass noun, we also see *de/ge3*:

(i)   a.   hou2-yam2      ge3     seoi2          (Cantonese)
           tasty          CL-u    water
           'tasty water'
      b.   māmā   de      tāng                   (Mandarin)
           mom    CL-u    soup
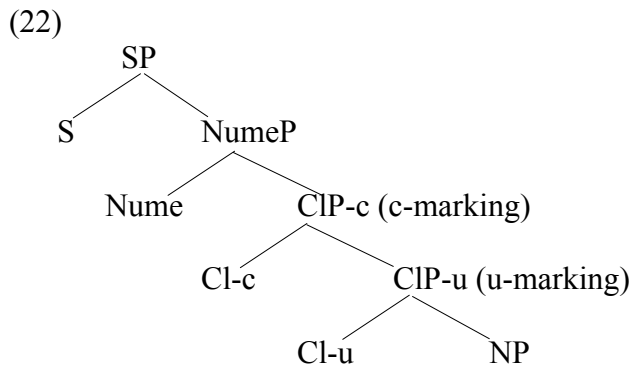           'mom's soup'

In these cases we assume that a portion of the mass (water, soup) is isolated after which the modifier is added. This is how we get the reading that not all the water or soup in the world is modified, but only a portion of it.

ungrammatical: it contains u-marker *ge3*, and it contains *ga3* (cf. *bún2* in (14a)). But *ga3* cannot be a u-marker, because we already have one, and it cannot be a c-marker either, because it is unaccompanied by a numeral. The sentence in (21a) contrasts the one in (21b), in which *ga3* is preceded by a numeral.[7]

(21) a.    *ga3    hou2   leng3   ge3     ce1 hai2 oi6min6
          CL-u    very    pretty   CL-u    car   be.at   outside
          'the pretty car is outside'
     b.        jau5     jat1-ga3   hou2 leng3   ge3   ce1   hai2 oi6min6
          there.is one-CL-c   very   pretty CL-u car   be.at outside
          'there is a pretty car outside'

We return to this below.

      Before continuing, we present the full structure of a Chinese DP we are assuming, including the layer hosting the demonstrative, dubbed "SP", short for "Specificity Phrase", as it is called in Sio (2006) (see also Cheng and Sybesma 1999; Sybesma and Sio 2008).

(22)

```
        SP
       /  \
      S    NumeP
           /   \
        Nume    ClP-c (c-marking)
                /    \
             Cl-c     ClP-u (u-marking)
                      /     \
                   Cl-u      NP
```

## 4. The distribution of Mandarin *de* and Cantonese *ge3*: some generalizations

In this section we take a closer look at the distribution of Mandarin *de* and Cantonese *ge3*. We start out from Mandarin. The data in (1) (repeated below as (23)) show modification of bare nouns. When we look at the pattern with respect to the optionality of *de*, we observe, as mentioned above, that *de* is optional only in (23a). In reference to Zhu's distinction between SAs and CAs, we could call all modifiers in (1b) to (1g) "complex", which then enables us to make the generalization that all complex modifiers need *de*, while simplex ones do not. This is basically the same as what Paul (2005) says.

(23) a.     dà (de) yú                           simple adjective
            big DE fish
            'big fish'
     b.     fēicháng       dà *(de) yú                complex adjective

---

[7] In (21b), *jat1* 'one' can be left phonologically empty, which does not affect the argument; see Cheng and Sybesma (1999).

extraordinarily big   DE fish
　　　　　　　'very big fish'
　　　c.　　Zhāng Sān *(de) yīfu　　　　　　　　possessor
　　　　　　　Zhang San DE clothes
　　　　　　　'Zhang San's clothes'
　　　d.　　méi  mǎi-guo  shū *(de) rén　　　　　　RC
　　　　　　　NEG buy-EXP  book DE person
　　　　　　　'people who have never bought a book'
　　　e.　　tā  chàng gē *(de) shēngyīn　　　　　gapless RC
　　　　　　　3S sing   song DE   voice
　　　　　　　'the voice with which he sings'
　　　f.　　duì　　　érzi *(de) tàidu　　　　　　PP
　　　　　　　regarding son   DE  attitude
　　　　　　　'the attitude towards his son'
　　　g.　　yǐqián *(de) zǒngtǒng　　　　　　non-predicative modifier
　　　　　　　former DE president
　　　　　　　'the former president'

Interestingly, when we look at phrases in which the modifier precedes the demonstrative, we see a pattern that is exactly the mirror image:

(24)　a.　　dà *(de) nèi-tiáo yú
　　　　　　　big  DE that-CL fish
　　　　　　　'that *big* fish'
　　　b.　　fēicháng　　　dà <sup>??</sup>(de) nèi-tiáo yú
　　　　　　　extraordinarily big   DE that-CL  fish
　　　　　　　'that *terribly big* fish'
　　　c.　　Zhāng Sān (de) nèi-jiàn yīfu
　　　　　　　Zhang San  DE  that-CL clothes
　　　　　　　'that piece of clothing of *Zhang San*'s'
　　　d.　　méi  mǎi-guo  shū (de) nèi-ge  rén[8]
　　　　　　　NEG buy-EXP  book DE that-CL person
　　　　　　　'that person who never bought a book'
　　　e.　　tā  chàng gē  (de) nèi-ge shēngyīn
　　　　　　　3S sing   song DE  that-CL voice
　　　　　　　'that voice with which he sings'
　　　f.　　duì　　　érzi <sup>??</sup>(de) nèi-zhǒng tàidu
　　　　　　　regarding son   DE that-type attitude
　　　　　　　'that kind of attitude towards one's son'
　　　g.　　yǐqián (de) nèi-ge zǒngtǒng
　　　　　　　former DE  that-CL president
　　　　　　　'that former president'

---

[8] In the case of *object* relatives, speakers of Mandarin strongly prefer *de*. We think that this is related to a processing problem, since the sentence will be the same as a non-relative in terms of word order. This is different from a subject relative, which always yields a different word order.

Obviously, the generalization stating that all complex modifiers need *de*, while simplex ones do not, cannot be upheld. However, when we take the nature of the modifiee into account as well, we observe that when simplex modifies simplex and when complex modifies complex, *de* is not obligatory; *de* is only obligatory, when simplex modifies complex or when complex modifies simplex. There is just one exception to this generalization: when complex modifiers precede complex phrases such as [Nume-Cl-N], *de* is obligatory (except in kinship cases, which we said we would leave out of the discussion).

(25)   a.   wǒ zuótiān   kàn-le   Zhāng Sān *(de) yī-běn shū
            1S  yesterday read-PERF Zhang San   DE one-CL book
            'yesterday I read one of Zhang San's books'
      b.   wǒ zuótiān   pèng-dào-le   nǐ qùnián gěi wǒ jièshào *(de) yī-ge   zuòjiā
            1S yesterday ran-into-PERF 2S last.year to 1S  introduce DE one-CL author
            'yesterday I ran into an author you introduced to me last year'

This means that complexity is not the (only) issue. Referential properties also play a role: [Nume-Cl-N] phrases are the only explicitly indefinite phrases in the language, and *de* is obligatory.

   In order to enable ourselves to explain the presence, absence and optionality of *de*, which we attempt to do in section 5, we need to explore a similar range of data in Cantonese. Cantonese shows basically the same pattern as we saw in Mandarin, except that, as Sio (2006, 150) reports (also, see Cheng 1998), *ge3* is quite strongly dispreferred in demonstrative cases (that is, the Cantonese counterparts of (24a)-(24g)), much more strongly than in Mandarin.[9]

   A more general point of difference between Cantonese and Mandarin is that in the latter we generally never find classifiers unpreceded by a numeral or a demonstrative, while such [Cl-N] phrases abound in the former; we already saw an example in (14a). What is relevant for us now, is that [Cl-N] phrases can also be modified, and that, when they are, we see that there are restrictions (see also Sio 2006). First, modifiers directly preceding the classifier never have *ge3*. Secondly, of all the different types of modifiers we have seen so far (in (1), (23)-(24)), not all can appear directly in front of the [Cl N] phrase; in fact, for reasons unclear to us, adjectives, complex and simplex alike, cannot immediately appear before [Cl-N] (with or without *ge3*) (we will have nothing to say about this in this paper).

(26)   a.   *daai6 (ge3) tiu4 jyu2
               big   GE   CL  fish
               intended: 'the big fish'
      b.   *hou2 daai6 (ge3) tiu4 jyu2

---

[9] It must be noted that the Cantonese counterpart of (24a) is okay, but only in contrastive contexts, in which case, however, the N is preferably dropped.

   More generally, Arsenijevic and Sio (2008) present data from Xiang in which the marker is ungrammatical with a demonstrative. As they note, the ban on the co-occurrence of the demonstrative and the marker seems to go from weak (Mandarin) to strong (Xiang), with Cantonese in the middle, but close to Xiang.

very big     GE   CL  fish
intended: 'the very big fish'
c.        zoeng1 saam1 (*ge3) gin6 saam1
          Zoeng Saam     GE   CL  clothes
          'Zoeng Saam's piece of clothing'
d.       mou5 maai5-gwo3 syu1 (*ge3) go3 yan4
          NEG   buy-EXP     book   GE   CL  person
          'the person who never bought a book'
e.        keoi5 coeng3 go2 (*ge3) baa2 seng1
          3S     sing     song  GE   CL  voice
          'the voice with which he sings'
f.        deoi3      zai2 (*ge3) go3 taai3dou6
          regarding son    GE   CL  attitude
          'the attitude towards son'
g.       ji5-cin4 (*ge3) go3 zung2tung2
          former     GE   CL  president
          'the former president'

What is interesting about this pattern, is that, this time, the CAs fall into the same category as the SAs, and not, as has been the case so far, with the other "complex" modifiers. Another interesting fact is that, as observed by Sio (2006, 67), [Mod Cl N] phrases are always definite. The question is why they *have to* be definite, since [Cl-N] phrases can otherwise be interpreted as definite and indefinite. This seems to confirm that differences in complexity cannot be the whole story and that definiteness also plays a role. Leaving this point, as well as the optionality issue aside for now, we observe on the basis of all data reviewed so far, that we don't find *de/ge3* with nouns that are explicitly marked as definite. This is our generalization (see also Sio 2006):

(27)     Phrases with *de/ge3* cannot modify definite nouns. They only modify indefinites.

Interpreting nominal phrases containing a demonstrative as definite (but see discussion in Sybesma and Sio 2008), this generalization is correct for Cantonese. We do however have to add that, although it may be true that phrases with *ge3* only modify indefinites, not all indefinites can be modified by modifiers involving *ge3*: indefinite [Cl-N] phrases cannot. For Mandarin, the generalization in (27) is only weakly true, in that *de* is only dispreferred right before a demonstrative, not completely barred.


## 5. Summary of the data and explanations

What we have observed so far can schematically be summarized as follows, where "X" is any bare modifier, and "X de/ge3" is any modifier with the marker (see also Sio 2006):

         **Mandarin and Cantonese**
(28)   a.      (Dem) Nume Cl {X de/ge3}/{X$_{SA}$} N
        b.      {X de/ge3} Nume Cl N

c.       {X (de/ge3)} Dem (Nume) Cl  N
         [in (28c): {X} (without *de/ge3*) preferred, strongly so in Cantonese; fn. 9]

         **Cantonese**
d.       Cl {X}/*{X ge3} N
e.       {X}/*{X ge3} Cl N [+def]
f.       *{X}/*{X ge3} Cl N [–def]

Let us now go back to the questions we posed in the beginning of this paper, given here once more, with *ge3* added in:

**I.** what is the distribution of modifiers with and those without *de/ge3* (and what are the meaning differences, if there is any)?
**II.** what are the structural positions of modifiers with *de/ge3* and what are the structural positions of those without?
**IIIa.** what is the internal make-up of modifiers with and those without *de/ge3*?
       which includes the separate question
**IIIb.** what is the nature of *de/ge3*? and
**IV.** how do these issues interrelate?

Some of these questions we are now in a position to answer. This certainly applies to I, which is answered in (28) (and, with respect to the meaning differences, in the discussion in the preceding sections).
        With respect to IIIb, we have put forth the hypothesis that *de/ge3* is a classifier performing the u-marking function. This hypothesis still needs to be fleshed out, which we will do to a certain extent while answering the remaining questions, II, IIIa and IV, also explaining the ungrammaticality of certain combinations.

**5.1. *de/ge3* or no *de/ge3***
The question posed in II is naturally related to IIIa. Let us now consider the answers to these questions under the assumption that *de/ge3* is a classifier (Cl-u). Consider first the structural distribution in (28a).
        In the case in which a bare N is modified by a simplex X, we assume that the X is combined with X at the lexical level. In other words, [A N] phrases may be considered compounds (see Paul 2005 and Feng 2001 for arguments against treating all [A N]'s as compounds). Consider now the case when [X de/ge3] modifies a bare N. Since *de/ge3* is in the head of ClP-u, we assume that the modifier is in the SpecClP.[10]  (In (28a), the classifier following the numeral is an instantiation of Cl-c.)  The difference between modification with and without *de/ge3* must, we think, be interpreted as what we saw in Thai. With *de/ge3*, we modify an individual, without *de/ge3* we modify the N before the individual has been singled out.[11]

---

[10] At this point, it is unclear whether there is a difference putting the modifier in the inner or outer Spec of ClP.
[11] For the sake of clarity, at the level of the bare noun, the phrase is still indefinite, and addition of *de/ge3* is unproblematic in view of the generalization in (27). The addition of a modifier to the NP, does, of course,

15

A question we need to address is why we need the *de*, a u-marker, in Mandarin. After all, we have been assuming that we need the u-marker in Cantonese as the result of a general rule that says that count nouns need to be marked as such. But for Mandarin we claimed that count nouns are already u-marked in the lexicon. So, what do we need *de* for? Our answer is that there are contexts in which the u-marking must be syntactically active, in which case the u-marking in the lexicon is doubled by a u-marking operation in syntax. Presumably, modification of individuals requires that the individual is syntactically active, requiring syntactic u-marking in such cases. In Mandarin, *de* is the overt realization of the Cl-u (see below for the condition in which Cl-u needs to be overtly realized).

Turning now to (28b) and (28c) in relation to question II, consider first (28c). The questions raised by (28c) are: (i) is there real optionality of *de/ge3*? (ii) why are modifiers with *de*/*ge3* dispreferred?

With respect to the second question, Arsenijevic and Sio (2007, 2008) explain the strong dispreference for *ge3* in Cantonese with reference to the deictic component of *ge3*: the demonstrative is already deictic; it is thus impossible to introduce another deictic element. The problem with this explanation is that even in Cantonese, *ge3* is not completely illicit preceding a demonstrative and that in Mandarin, the dispreference of the construction with *de* is much less strong than is the case in Cantonese with *ge3*.

To answer the question concerning optionality, recall that when the modifier precedes the Dem-Nume-Cl-N sequence, it acquires a contrastive focus reading (see examples (2a,b) vs. (3a,b)). We propose that there are two structures which are compatible with the cases in which the modifier precedes the Dem-Nume-Cl-N sequence. First, taking into consideration that a contrastive focus reading is actually present in such cases, we suggest that the modifier is moved to the left-periphery of the DP for focus (see Aboh 2004 among others for the left-periphery of DP structures). Assuming that Cl-u need not (or cannot) be overtly spelled out when no XP is in its Spec, we derive the case in which the modifier precedes the Dem-Nume-Cl-N sequence without *de/ge3*. Aside from the structure associated with contrastive focus, we suggest that (28c) with the presence of *de/ge3* involves an appositive structure. That is, (28c) has the following structure:

(29)  [$_{DP}$ X *de/ge3* NP] [$_{DP}$ Dem (Nume) Cl NP]

(29) is the structure for typical noun phrase apposition such as [John Smith, the President]. The only difference is that in (29), the first DP involves an empty head noun, just as it is possible to have *the big one* in English (with *one* as an empty noun). We assume that this is also the case in (28b), involving indefinite noun phrases in apposition (though in the case of appositives, it is probably a specific indefinite).[12]

**5.2. Cantonese**
Turning to (28d), we need to explain why [Cl X-ge3 N] is ungrammatical. The explanation has already been given above (cf. the discussion on (21) and fn. 6). Because the Cl here is unaccompanied by a numeral, it must be a u-marker. This means that we

---

not make any difference for the possibility of building more structure, involving a numeral and, ultimately, a demonstrative.

[12] This implies that even in cases with apposition, there is a contrastive reading.

have two syntactic u-markers in one phrase, which is illicit. The same explanation can be extended to [X-ge3 Cl N] in (28e): here too, the Cl is unaccompanied by a numeral and must hence be a u-marker. The addition of another u-marker (*ge3*) leads to ungrammaticality. Such addition is also not necessary, since the modifier can be merged to the Cl-u already present (see also fn. 13).

Finally, (28f) raises another two questions: (i) why can [Mod Cl N] phrases never be indefinite, while [Cl N] phrases can be both definite and indefinite? (ii) why can't the indefinite [Cl N] phrase not be modified by a modifier with *de/ge3* (this question is especially relevant in view of our generalization in (27)). Here we may adopt Sio's (2006) technical explanation. Following Cheng and Sybesma (1999) in assuming that indefinite [Cl N] phrases involve a phonologically empty NumeP, Sio proposes that adjunction to a phrase with an empty head is not possible, thus explaining why indefinite [Cl N] phrases cannot be modified.

Summing up the distribution of *ge3* in Cantonese, it should be noted that the Cl-u can host either *ge3* or a classifier in Cantonese, since Cantonese classifiers can function as Cl-u's. Therefore, a modifier (possessive NP, relative clauses) can merge with ClP-u, with Cl-u being spelled out either as a classifier or as *ge3*.[13]

## 5.3. Remaining cases and issues

One issue that immediately jumps out in relation to relative clauses is that, in our analysis, *de/ge3* does not really form a constituent with the relative clause. Consider first whether our treatment of *de/ge3* as Cl-u is problematic for relative clauses. There are two distributional patterns we need to consider:

(30)  a.  Dem-Nume-CL-[Rel clause]-*de/ge3*-N
      b.  [Rel clause]-(*de/ge3*)-Dem-Nume-Cl-N

These two cases are actually already discussed in relation to (28a) and (28c): (30a) is a typical case of a restrictive relative clause, in which case the relative clause is generally considered to be adjoined to the NP, and in our case a ClP-u, the first FP on top of NP. This means that *de/ge3* is not part of the relative clause. Considering (30b), we suggest that the relative clause is on a par with possessors in that it can be in the left periphery of the DP to mark focus. When this happens, *de/ge3* is not present. When *de/ge3* is present, just as we have in the case of possessors, we have an appositive structure, with an empty noun, as in (29).[14]

## 6. Conclusion

---

[13] There is one more issue here, viz., the fact, not mentioned so far, that [Nume-Cl *de/ge3* N] is ungrammatical (e.g., Cantonese *jat1 ga3 ge3 ce1* /one-CL GE3 car/). It is possible that Cl-u can only be realised as *de/ge3* if the SpecCl-u is filled by a modifier. With "massifiers" this sequence is possible (e.g., Mandarin sān xiāng de shū /three box DE book/ 'three boxfuls of books'), but the derivation is different; see Cheng and Sybesma (1998).

[14] Whether or not Chinese relative clauses yield both restrictive and non-restrictive readings has been a controversial issue. The appositive structure that we are suggesting here implies that it is also possible to have an appositive reading. For discussion, see Lin (2004).

We conclude that treating the "modification markers" *de* and *ge3* as u-markers, filling the Cl-u position in the F-structure of the noun phrase, enables us to explain the most important distributional properties of these elements. It also enables us to have a simple generalization about what licenses elliptical noun phrases: much in line with the findings by others, we conclude that it is always a classifier-type of element which licenses NPE. Furthermore, in relation with this, we have seen that modifiers of every type are always in a specifier position: of N/NP, ClP-u or of SP.

This basiscally answers question IV: the assumptions (i) that *de* and *ge3* are u-markers; (ii) that we only find one syntactic u-marker in the F-domain of an NP; and (iii) that modifiers are in the Spec of some projection in the phrase, together give us the distributional facts in (28).

## References

Aboh, Enoch. 2004. *The morphosyntax of complement-head sequences, clause structure and word order patterns in Kwa*. Oxford: Oxford University Press.

Alexiadou, Artemis, and Gengel, Kirsten. 2008. NP ellipsis without focus movement/projections: the role of Classifiers. Ms. Universität Stuttgart.

Arsenijevic, Boban, and Sio, Joanna. 2007. Talking about Classifiers, the Cantonese *ge* is a curious one. Paper presented in WSM, Leiden University.

Arsenijevic, Boban, and Sio, Joanna Ut-Seong. 2008. The Cantonese *ge*. Paper presented at ARF 2008, University of Hong Kong.

Borer, Hagit. 2005. *In name only (Structuring sense I)*. Oxford: Oxford University Press.

Chao, Yuen Ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Cheng, Lisa L.-S. 1986. *de* in Mandarin. *Canadian Journal of Linguistics* 31:313-326.

Cheng, Lisa L.-S. 1998. Marking modification in Cantonese and Mandarin. Paper presented at SOAS, London.

Cheng, Lisa L.-S. 2009. On *every* type of quantificational expression in Chinese. In *Quantification, definiteness, and nominalization*, eds. Monika Rather and Anastasia Giannakidou, 53-75. Oxford: Oxford University Press.

Cheng, Lisa L.-S., and Sybesma, Rint. 1998. *Yi-wan tang, yi-ge Tang*: Classifiers and massifiers. *Tsinghua Journal of Chinese Studies*. New Series Vol. 28, 385-412.

Cheng, Lisa L.-S., and Sybesma, Rint. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30:509-542.

Cheng, Lisa L.-S., and Sybesma, Rint. 2006. A Chinese relative. *Organizing grammar: Linguistic studies in honor of Henk van Riemsdijk*, ed. N. Corver, H. Broekhuis, R. Huybregts, U. Kleinhenz, J. Koster, 69-76. Berlin: Mouton.

Chierchia, Gennaro. 1998. Reference to kinds across languages. *Natural Language Semantics* 6:339-405.

Croft, William. 1994. Semantic universals in classifier systems. *Word* 45:145-171.

den Dikken, Marcel, and Singhapreecha, Pornsiri. 2004. Complex noun phrases and linkers. *Syntax* 7:1-54.

Doetjes, Jenny. 1997. *Quantifiers and selection: on the distribution of quantifying expressions in French, Dutch and English*. Doctoral dissertation, Leiden University.

Féng, Shēnglì. 2001. Lùn Hànyǔ cí de duō-wéi-xìng [The multidimensional properties of Chinese words]. *Dāngdài Yǔyánxué* 3/3:161–174.

Fu, Jingqi. 1987. La structure du syntagme nominal en chinois. Doctoral dissertation, Paris.

Fung, Roxanne S.-Y. 2000. *Final particles in Standard Cantonese: Semantic extension and pragmatic inference.* Doctoral dissertation, Ohio State University.

Huang, Shi-Zhe. 2006. Property theory, adjectives and modification in Chinese. *Journal of East Asian Linguistics* 15:343-369.

Iljic, Robert. 1994. Quantification in Mandarin Chinese: two markers of plurality. *Linguistics* 32:91-116.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, Mass: MIT Press.

Kookiattikoon, Supath. 2001. *The syntax of classifiers in Thai.* Doctoral dissertation, University of Kansas.

Li, Charles N., and Thompson, Sandra A. 1981. *Mandarin Chinese. A functional reference grammar.* Berkeley: University of California Press.

Lin, Jo-wang. 2004. On restrictive and non-restrictive relative clauses in Mandarin Chinese, *Tsinghua Journal of Chinese Studies. New Series* 33/1, 199-240.

Paris, Marie-Claude. 1979. *Nominalization in Mandarin Chinese. The morpheme 'de' and the 'shi … de' constructions.* Université Paris 7.

Paul, Waltraud. 2005. Adjectival modification in Mandarin Chinese and related issues. *Linguistics* 43:757-793.

Picallo, Carmen. 2006. On gender and number. Ms. Universitat Autònoma de Barcelona.

Pirani, Laura. 2007. *Morphological processes in the Mandarin nominal domain.* Doctoral dissertation, Universita' degli studi di Verona.

Rubin, Edward. 2003. The structure of modifiers. Ms. University of Utah.

Sharvy, Richard (1978). Maybe English has no count nouns: notes on Chinese semantics. An essay in metaphysics and linguistics. *Studies in language* 2/3, 345-365.

Simpson, Andrew. 2002. On the status of 'modifying' DE and the structure of the Chinese DP. *On the formal way to Chinese languages*, eds. Sze-Wing Tang and Chen-Sheng Liu. CSLI.

Sio, Joanna Ut-seong. 2006. *Modification and reference in the Chinese nominal*. Doctoral dissertation, Leiden University.

Sybesma, Rint. 2007. Běifāng fāngyán hé Yuèyǔ zhōng míngcí de kěshǔbiāojì (Markers of countability on the noun in Mandarin and Cantonese)". *Yǔyánxué lùncōng* 35, 234-245.

Sybesma, Rint. 2008. Classifiers, number and countability. Ms. Leiden University.

Sybesma, Rint and Sio, Joanna. 2008. D is for Demonstrative. Investigating the position of the demonstrative in Chinese and Zhuang. *The Linguistic Review* 25/3-4 [Special issue on Syntactic categories and their interpretation in Chinese, ed. Huba Bartos], 453-478.

Xu, Zheng (2006). On the formation of Adjective-Noun combinations in Mandarin Chinese. *NACCL 17 Proceedings*, ed. Gao Qian.

Zhū, Déxī. 1956. Xiàndài Hànyǔ xíngróngcí yánjiū [A study of adjectives in Modern Chinese]. *Zhōngguó Yǔwén* 5:447-458.

Zhū, Déxī. 1961. Shuō *de* [On *de*]. *Zhōngguó Yǔwén* 1:1-15.